

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification <sup>6</sup> :	A1	(11) International Publication Number:	WO 99/02989
G01N 33/50, C12Q 1/02		(43) International Publication Date:	21 January 1999 (21.01.99)

(21) International Application Number:	PCT/GB98/02042		
(22) International Filing Date:	10 July 1998 (10.07.98)		
(30) Priority Data:	9714582.5	10 July 1997 (10.07.97)	GB
	9718261.2	28 August 1997 (28.08.97)	GB
(71) Applicant ( <i>for all designated States except US</i> ):	MEDICAL RESEARCH COUNCIL [GB/GB]; 20 Park Crescent, London W1N 4AL (GB).		
(72) Inventors; and			
(75) Inventors/Applicants ( <i>for US only</i> ):	FERSHT, Alan [GB/GB]; Cambridge University Chemical Laboratory, Lensfield Road, Cambridge CB2 1EW (GB). CHATELLIER, Jean [FR/GB]; MRC Laboratory of Molecular Biology, Medical Research Council Centre, Hills Road, Cambridge CB2 2QH (GB).		
(74) Agents:	MASCHIO, Antonio et al.; D Young & Co., 21 New Fetter Lane, London EC4A 1DA (GB).		

(54) Title: CHAPERONE FRAGMENTS

(57) Abstract

The invention relates to a method for determining whether a fragment of a molecular chaperone is active *in vivo*, comprising the steps of: a) providing a cell with a deficient molecular chaperone activity; b) administering the molecular chaperone fragment to the cell; and c) determining whether the molecular chaperone fragment complements the deficient endogenous molecular chaperone activity, as well as to methods for providing molecular chaperone activities *in vivo*.

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		

### Chaperone Fragments

The present invention relates to chaperone polypeptides which are active in folding or maintaining the structural integrity of other proteins. In particular, the present 5 invention relates to fragments of chaperone polypeptides which are active *in vivo* and can complement or potentiate the activity of intact molecular chaperones *in vivo*, and to methods for identifying such fragments.

Chaperones are in general known to be large multisubunit protein assemblies essential 10 in mediating polypeptide chain folding in a variety of cellular compartments. Families of chaperones have been identified, for example the chaperonin hsp60 family otherwise known as the cpn60 class of proteins are expressed constitutively and there are examples to be found in the bacterial cytoplasm (GroEL), in endosymbiotically derived mitochondria (hsp60) and in chloroplasts (Rubisco binding protein). Another chaperone 15 family is designated TF55/TCP1 and found in the thermophilic archaea and the evolutionarily connected eukaryotic cytosol. A comparison of amino acid sequence data has shown that there is at least 50% sequence identity between chaperones found in prokaryotes, mitochondria and chloroplasts (Ellis R J and Van der Vies S M (1991) *Ann Rev Biochem* **60**: 321-347).

20 A typical chaperonin is GroEL which is a member of the hsp60 family of heat shock proteins. GroEL is a tetradecamer wherein each monomeric subunit (cpn60m) has a molecular weight of approximately 57kD. The tetradecamer facilitates the *in vitro* folding of a number of proteins which would otherwise misfold or aggregate and precipitate. The structure of GroEL from *E. coli* has been established through X-ray 25 crystallographic studies as reported by Braig K *et al* (1994) *Nature* **371**: 578-586. The holo protein is cylindrical, consisting of two seven-membered rings that form a large central cavity which according to Ellis R J and Hartl F U (1996) *FASEB Journal* **10**: 20-26 is generally considered to be essential for activity. Some small proteins have 30 been demonstrated to fold from their denatured states when bound to GroEL (Gray T E and Fersht A R (1993) *J Mol Biol* **232**: 1197-1207; Hunt J F *et al* (1996) *Nature* **379**: 37-45; Weissman J S *et al* (1996) *Cell* **84**: 481-490; Mayhew M *et al* (1996) *Nature*

379: 420-426; Corrales F J and Fersht A R (1995) Proc Nat Acad Sci 92: 5326-5330) and it has been argued that a cage-like structure is necessary to sequester partly folded or assembled proteins (Ellis R J and Hartl F U (1996) *supra*.

5 The entire amino acid sequence of *E. coli* GroEL is also known (see Braig K *et al* (1994) *supra*) and three domains have been ascribed to each cpn60m of the holo chaperonin (tetradecamer). These are the intermediate (amino acid residues 1-5, 134-190, 377-408 and 524-548), equatorial (residues 6-133 and 409-523) and apical (residues 191-376) domains.

10 International patent application WO98/13496 (incorporated herein by reference) describes the generation of fragments of molecular chaperones which surprisingly possess molecular chaperone activity in monomeric form. A number of preferred fragments are identified, some of which are based on the apical domain of GroEL. One 15 of the most preferred fragments includes amino acid residues 191-376 of *E. coli* GroEL. However, the activity reported in WO98/13496 is largely *in vitro* activity. There are no examples of *in vivo* assessment of chaperone activity.

### Summary of the Invention

20 It has surprisingly been found that *in vitro* activity of molecular chaperone fragments does not necessarily reflect *in vivo* activity of the same fragments. Accordingly, in a first aspect, the present invention provides a method for determining whether a fragment of a molecular chaperone is active *in vivo*, comprising the steps of:

- 25 a) providing a cell with a deficient molecular chaperone activity;  
b) administering the molecular chaperone fragment to the cell;  
c) determining whether the molecular chaperone fragment complements the deficient endogenous molecular chaperone activity.

In a further aspect, the invention provides a method for providing a chaperone activity *in vivo*, comprising administering to a cell a fragment of a molecular chaperone which has *in vivo* activity.

- 5 Advantageously, the molecular chaperone fragment complements a mutant or depressed endogenous molecular chaperone activity.

According to a third aspect, the invention relates to the use of a fragment of a molecular chaperone to complement a mutant or deficient molecular chaperone activity  
10 *in vivo*.

#### Brief Description of the Drawings

**Fig. 1.** Schematic representation of the plasmid construction and organisation. amp, β-lactamase; cm, chloramphenicol acetyl transferase; R, resistance; rbs, ribosome binding site; p/o, promoter/operator: tet, tetracycline. Relevant restriction sites are indicated.

**Fig. 2.** Three-dimensional structure of minichaperone GroEL(191-345) solved at 2.5 Å  
20 (Zahn, R., Buckle, A. M., Perret, S., Johnson, C. M. J., Corrales, F. J., Golbik, R. & Fersht, A. R. (1996) *Proc. Natl. Acad. Sci. USA* **93**, 15024-15029). Secondary structure representation is drawn with MolScript (Kraulis, P. (1991) *J. Appl. Crystallogr.* **24**, 946-950). Positions mentioned in the text are indicated (residues numbered as in Hemmingsen *et al.*, (1988) *Nature* **333**, 330-334).

25 **Fig. 3.** *In vitro* refolding of rhodanese in the presence of GroEL minichaperones. Relative enzymatic activity of rhodanese (0.1 μM) after refolding in the presence (+) or absence (-) of GroEL (2.5 μM monomer), GroES (2.5 μM monomer), ATP (2 mM), sht-GroEL191-376 (2.5 μM), sht-GroEL191-345 (2.5 μM), sht-GroEL193-335 wild-type and mutant Y203E (2.5 μM), or bovine serum albumin (BSA; 45 μg/ml), from 8  
30

M urea (U). The yield of refolding activity was measured after 15 min at 25°C. 100% activity was obtained with native rhodanese (N). Standard error bars are shown.

Fig. 4. Suppression of the temperature-sensitive phenotypes of *E. coli* SV2 (*groEL44*)  
5 and SV6 (*groEL673*) strains by over-expression of GroEL or minichaperones. The  
vectors are indicated on the right-hand side.

Fig. 5. Supplementation of low levels of GroEL in *E. coli* by minichaperones. *E. coli*  
AI90[pBAD-EL] has the chromosomal *groEL* gene deleted by P1 transduction (Ivic,  
10 A., Olden, D., Wallington, E. J. & Lund, P. A. (1997) *Gene* 194, 1-8) and essential  
GroEL is provided by a plasmid-borne copy of the gene pBAD-EL, which can be  
tightly regulated by arabinose (Guzman, L.-M., Belin, D., Carson, M. J. & Beckwith,  
J. (1995) *J. Bacteriol.* 177, 4121-4130). In the absence of arabinose, the cells are not  
viable, unless pJCGroEL(1-548) is present, which expresses intact GroEL (O). pJCsht  
15 (Δ) and pJCGroELsht-(193-335)Y203E (□) plasmids are controls, and cells containing  
these show increased viability with increasing arabinose concentrations. Over-  
expression of sht-GroEL(191-376) (●) marginally increases viability; sht-GroEL(191-  
345) (▲) and (193-335) (■) minichaperones both give significant increases. Each value  
recorded is the plating efficiency relative to that in the presence of 0.2% arabinose  
20 (100%).

Fig. 6. Effect of over-expressing GroEL or the minichaperones on Lorist6 replication.  
Cultures of TG1 carrying Lorist6 and one of the pBAD30 series were titred on LB  
25 plates containing kanamycin and ampicillin in presence or absence of 0.2% arabinose  
at 37°C. The percentage of cells forming kanamycin resistant colonies in presence of  
arabinose compared to the number formed in the absence of arabinose (100%) is shown  
for each of the pBAD30 series of expression vectors. Lorist6 plasmid encodes  
resistance to kanamycin and uses the bacteriophage λ origin of replication (Gibson, T.  
J., Rosenthal, A. & Waterston, R. H. (1987) *Gene* 53, 283-286). Arabinose induces  
30 expression of the pBAD30 series vectors; loss of kanamycin resistance reflects  
inhibition of Lorist6 replication.

**Detailed Description of the Invention**

- 5    In accordance with the present invention, it has surprisingly been found that although various fragments of molecular chaperones are capable of promoting the folding of polypeptides *in vitro*, this does not necessarily mean that they are capable of providing some or all of the functions of endogenous molecular chaperones *in vivo*. This conclusion is of great importance, because in many therapeutic applications envisaged  
10    for molecular chaperones, which involve the correction of defects in protein folding, an *in vivo* activity is required. Thus, the present invention provides a method for identifying fragments of molecular chaperones which are potentially useful in *in vivo* applications, as well as methods for using such fragments *in vivo*.
- 15    A further surprising finding of the present invention is that fragments of molecular chaperones are able to replace specific activities of molecular chaperones *in vivo*, but often not able to replace the chaperone activity in its entirety. This means that, where the molecular chaperone is essential for cell viability, it is not possible to test the activity of chaperone fragments by deleting the endogenous chaperone activity and  
20    attempting to rescue it. Thus, the invention provides for the use of deficiencies in molecular chaperones in order to screen chaperone fragments for the ability to complement these deficiencies.

In the present text, references to the singular, for example as in "a cell", are to be  
25    interpreted as references to the singular both in isolation and when encompassed in a plurality. Thus, terms such as, for example, "a cell" and "one or more cells" and "at least one cell" are equivalent.

As used herein, a fragment of a molecular chaperone is any fragment of a molecular  
30    chaperone polypeptide which has molecular chaperone activity. Molecular chaperones are well known in the art, several families thereof being characterised. The invention

is applicable to any molecular chaperone molecule, which term includes, for example, the molecular chaperones selected from the following non-exhaustive group:

p90 Calnexin	Salopek <i>et al.</i> , J. Investig Dermatol Symp Proc (1996) 1:195
HSP family	Walsh <i>et al.</i> , Cell Mol. Life Sci. (1997) 53:198
HSP 70 family	Rokutan <i>et al.</i> , J. Med. Invest. (1998) 44:137
DNA K	Rudiger <i>et al.</i> , Nat. Struct. Biol. (1997) 4:342
DNAJ	Cheetham <i>et al.</i> , Cell Stress Chaperones (1998) 3:28
HSP 60 family; GroEL	Richardson <i>et al.</i> , Trends Biochem. (1998) 23:138
ER-associated chaperones	Kim <i>et al.</i> , Endocr Rev (1998) 19:173
HSP 90	Smith, Biol. Chem. (1988) 379:283
Hsc 70	Hohfeld, Biol. Chem. (1988) 379:269
sHsps; SecA; SecB	Beissinger <i>et al.</i> , Biol. Chem. (1988) 379:245
Trigger factor	Wang <i>et al.</i> , FEBS Lett. (1998) 425:382
zebrafish hsp 47, 70 and 90	Krone <i>et al.</i> , Biochem. Cell Biol. (1997) 75:487
HSP 47	Nagata, Matrix Biol. (1998) 16:379
GRP 94	Nicchitta <i>et al.</i> , Curr. Opin. Immunol. (1998) 10:103
Cpn 10	Cavanagh, Rev. Reprod. (1996) 1:28
BiP	Sommer <i>et al.</i> , FASEB J. (1997) 11:1227
GRP 78	Brostrom <i>et al.</i> , Prog. Nucl. Acid. res. Mol. Biol. (1998) 58:79
C1p, FtsH	Suzuki <i>et al.</i> , Trends Biochem. Sci. (1997) 22:118
Ig invariant chain	Weenink <i>et al.</i> Immunol. Cell biol. (1997) 75:69
mitochondrial hsp 70	Horst <i>et al.</i> , BBA (1997) 1318:71
EBP	Hinek, Arch. Immunol. Ther. Exp. (1997) 45:15
mitochondrial m-AAA	Langer <i>et al.</i> , Experientia (1996) 52:1069
Yeast Ydj1	Lyman <i>et al.</i> , Experientia (1996) 52:1042
Hsp 104	Tuite <i>et al.</i> , Trends Genet. (1996) 12:467
ApoE	Blain <i>et al.</i> , Presse Med. (1996) 25:763

Syc	Wattiau <i>et al.</i> , Mol. Microbiol. (1996) 20:255
Hip	Ziegelhoffer <i>et al.</i> , Curr. Biol. (1996) 6:272
TriC family	Hendrick <i>et al.</i> , FASEB J. (1995) 9:1559
CCT	Kubota <i>et al.</i> , Eur. J. Biochem. (1995) 230:3
PapD, calmodulin	Stanfield <i>et al.</i> , Curr. Opin. Struct. Biol. (1995) 5:103

Preferably, the molecular chaperone is a chaperone of the hsp60 class, such as the *E. coli* chaperone GroEL.

- 5     The method of the invention allows molecular chaperone fragments active *in vivo* to be identified empirically. Accordingly, in one embodiment, fragments of molecular chaperones may be generated randomly and tested by the method of the invention to assess *in vivo* activity.
- 10    The size of molecular chaperone fragments does not necessarily correlate with *in vivo* activity. Thus, the fragments used in the method of the invention may be substantially any size. Preferably, the minimum size of suitable fragments can be determined experimentally, for example by generating crystal structures of the fragments. residues which display no electron density, and are thus outside the folded structure of the
- 15    chaperone fragment, are less likely to be essential for chaperone activity. Conversely, fragments which do not possess enough sequence to form a stable folded structure at all are unlikely to be active as chaperones.

Preferably, fragments selected for analysis are of the order of between about 100 to 500 amino acids in size, advantageously about 120 to 400, more advantageously about 130 to 300 and most advantageously about 145 to 200 amino acids in length.

Preferably, however, fragments of hsp60 class molecular chaperones may be generated as set forth in WO98/13496. In particular, molecular chaperone fragments may be taken from, or comprise, residues 191-376 of GroEL or the equivalent amino acid positions in other hsp60 molecules.

Especially preferred are fragments 191-345 and 193-335 of GroEL, or equivalents thereof. 193-335 is the smallest GroEL fragment which has been demonstrated to possess chaperone activity *in vitro*. Other fragments, including 191-345 and 191-376, 5 possess *in vitro* chaperone activity. However, it has surprisingly been shown that although 191-376 has only very limited activity *in vivo* and is not able to significantly complement a deficiency in molecular chaperone activity, both of the smaller fragments 191-345 and 193-335 are active *in vivo* and can complement a deficiency in molecular chaperone activity.

10

The present invention is concerned with *in vivo* molecular chaperone activity. As used herein, *in vivo* describes an activity which takes place inside a living cell. It is not necessarily intended to refer to activity in live animals, but also to include activity in tissue and cell culture. "Activity" is any activity which is ascribable to a molecular chaperone. Typically, these activities include the facilitation of protein folding, the latter term including the folding of unfolded proteins and the maintenance of folded proteins in a correctly folded conformation. Other activities of molecular chaperones which are included are antiviral activities, protein transport activities and the like.

20

Preferably, *in vivo* activity is measured as the ability to rescue cell viability when chaperone activity is deficient. Molecular chaperones of the hsp60 such as GroEL are essential, and if they are ablated from a cell the cell will die. However, mutants may be designed in which the endogenous holochaperone is mutated, for example such that it is temperature sensitive. In this case, the holochaperone will be deficient at non-permissive temperatures. Alternatively, a cell may be transfected with a vector encoding the holochaperone or a fully functional variant thereof under the control of regulatable sequences, such that in response to the appropriate stimuli the chaperone will be produced at deficient levels. In systems such as these, cell growth is compromised in the presence of chaperone deficiency. The *in vivo* activity of a chaperone fragment may be defined as the ability of the fragment to restore all or part 25 of the chaperone activity to such cells, when the fragment is introduced therein.

30

A "deficient" chaperone activity, as referred to herein, is a chaperone activity which is not equal to that of a wild type molecular chaperone. For example, it may be quantitatively deficient, such that an insufficient amount of chaperone protein is 5 available in the cell. Alternatively, it may be qualitatively deficient, such that the chaperone is not able to mediate certain reactions, or cannot do so at the rate at which a wild type chaperone mediates them. Typically, when referring to a deficient chaperone activity, it is intended to refer to a deficiency in endogenous holochaperone activity, such as a deficiency in endogenous GroEL activity in *E. coli*. However, the invention 10 also allows for the endogenous activity to be replaced or augmented by an exogenous activity, which may then be rendered deficient. For example, the endogenous chaperone activity may be deleted and "rescued" with a regulatable chaperone activity encoded on a plasmid, which is susceptible to downregulation to a deficient state.

15 The cell may have a permanently deficient chaperone activity, in which case the deficiency is preferably not lethal, allowing the cells to grow in the absence of a functional chaperone fragment. Alternatively, the deficiency may be regulatable, for example by means of a temperature-sensitive chaperone mutant or the introduction of a regulatable chaperone coding sequence as set forth above.

20 The chaperone fragment need not be able to replace or complement all of the lost molecular chaperone activities. Indeed, it is believed that chaperone fragments such as GroEL 193-335 are only capable of complementing certain protein folding activities of holo GroEL *in vivo*. Whilst not wishing to be bound by theory, it is possible that the 25 complementing of these activities is itself sufficient to restore partial cell growth to a cell deficient in holo GroEL, for example by allowing holo GroEL to be diverted to other processes where the 193-335 fragment cannot replace it.

Cells suitable for practising the present invention may be of any suitable type. Host 30 cells such as prokaryote, yeast and higher eukaryote cells may be used. Suitable prokaryotes include eubacteria, such as Gram-negative or Gram-positive organisms,

such as *E. coli*, e.g. *E. coli* K-12 strains, DH5a and HB101, or Bacilli. Further hosts suitable for chaperone fragment encoding vectors include eukaryotic microbes such as filamentous fungi or yeast, e.g. *Saccharomyces cerevisiae*. Higher eukaryotic cells include insect and vertebrate cells, particularly mammalian cells, including human cells, 5 or nucleated cells from other multicellular organisms.

The propagation of vertebrate cells in culture (tissue culture) is a routine procedure. Examples of useful mammalian host cell lines are epithelial or fibroblastic cell lines such as Chinese hamster ovary (CHO) cells, NIH 3T3 cells, HeLa cells or 293T cells. 10 The host cells referred to in this disclosure comprise cells in in vitro culture as well as cells that are within a host animal.

Host cells are transfected with vectors according to the invention and cultured in conventional nutrient media modified as appropriate for inducing promoters, selecting 15 transformants, or amplifying the genes encoding the desired sequences. Heterologous DNA may be introduced into host cells by any method known in the art, such as transfection with a vector encoding a heterologous DNA by the calcium phosphate coprecipitation technique or by electroporation. Numerous methods of transfection are known to the skilled worker in the field. Successful transfection is generally recognised 20 when any indication of the operation of this vector occurs in the host cell. Transformation is achieved using standard techniques appropriate to the particular host cells used.

Incorporation of cloned DNA into a suitable expression vector, transfection of 25 eukaryotic cells with a plasmid vector or a combination of plasmid vectors, each encoding one or more distinct genes or with linear DNA, and selection of transfected cells are well known in the art (see, e.g. Sambrook et al. (1989) Molecular Cloning: A Laboratory Manual, Second Edition, Cold Spring Harbor Laboratory Press).

30 Transfected or transformed cells are cultured using media and culturing methods known in the art, preferably under conditions, whereby chaperone fragment encoded by the

DNA is expressed. The composition of suitable media is known to those in the art, so that they can be readily prepared. Suitable culturing media are also commercially available.

- 5 In accordance with the present invention, the molecular chaperone fragment being tested is administered to the test cell, for example, by administration of the fragment in polypeptide form, or by introduction into the cell of a nucleic acid encoding the fragment such that it is expressed in the cell. Insertion of nucleic acid sequences, by processes such as transfection, transduction, electroporation or otherwise, is preferred.

10

Preferably, nucleic acid encoding a molecular chaperone fragment is introduced by means of a vector.

As used herein, vector (or plasmid) refers to discrete elements that are used to 15 introduce heterologous DNA into cells for expression thereof. Selection and use of such vehicles are well within the skill of the artisan. Many vectors are available, and selection of appropriate vector will depend on, for example, the size of the DNA to be inserted into the vector, and the host cell to be transformed with the vector. Each vector contains various components depending on its function and the host cell for which it is 20 compatible. The vector components generally include, but are not limited to, one or more of the following: an origin of replication, one or more marker genes, an enhancer element, a promoter, a transcription termination sequence and a signal sequence.

Both expression and cloning vectors generally contain nucleic acid sequences that 25 enable the vector to replicate in one or more selected host cells. Typically in cloning vectors, this sequence is one that enables the vector to replicate independently of the host chromosomal DNA, and includes origins of replication or autonomously replicating sequences. Such sequences are well known for a variety of bacteria, yeast and viruses. The origin of replication from the plasmid pBR322 is suitable for most 30 Gram-negative bacteria, the  $2\mu$  plasmid origin is suitable for yeast, and various viral origins (e.g. SV 40, polyoma, adenovirus) are useful for cloning vectors in mammalian

cells. Generally, the origin of replication component is not needed for mammalian expression vectors unless these are used in mammalian cells competent for high level DNA replication, such as COS cells.

- 5    Most expression vectors are shuttle vectors, i.e. they are capable of replication in at least one class of organisms but can be transfected into another class of organisms for expression. For example, a vector may be cloned in *E. coli* and then the same vector may be transfected into yeast or mammalian cells even though it is not capable of replicating independently of the host cell chromosome. DNA may also be amplified by  
10   PCR and may be directly transfected into the host cells without any replication component.

Advantageously, a vector may contain a selection gene also referred to as selectable marker. This gene encodes a protein necessary for the survival or growth of  
15   transformed host cells grown in a selective culture medium. Host cells not transformed with the vector containing the selection gene will not survive in the culture medium. Typical selection genes encode proteins that confer resistance to antibiotics and other toxins, e.g. ampicillin, neomycin, methotrexate or tetracycline, complement auxotrophic deficiencies, or supply critical nutrients not available from complex media.  
20

As to a selective gene marker appropriate for yeast, any marker gene can be used which facilitates the selection for transformants due to the phenotypic expression of the marker gene. Suitable markers for yeast are, for example, those conferring resistance to antibiotics G418, hygromycin or bleomycin, or provide for prototrophy in an  
25   auxotrophic yeast mutant, for example the URA3, LEU2, LYS2, TRP1, or HIS3 gene.

Since the replication of vectors is conveniently done in *E. coli*, an *E. coli* genetic marker and an *E. coli* origin of replication are advantageously included. These can be obtained from *E. coli* plasmids, such as pBR322, Bluescript<sup>®</sup> vector or a pUC plasmid,  
30   e.g. pUC18 or pUC19, which contain both *E. coli* replication origin and *E. coli* genetic marker conferring resistance to antibiotics, such as ampicillin.

Suitable selectable markers for mammalian cells are those that enable the identification of cells competent to take up chaperone fragment nucleic acid, such as dihydrofolate reductase (DHFR, methotrexate resistance), thymidine kinase, or genes conferring 5 resistance to G418 or hygromycin. The mammalian cell transformants are placed under selection pressure which only those transformants which have taken up and are expressing the marker are uniquely adapted to survive. In the case of a DHFR or glutamine synthase (GS) marker, selection pressure can be imposed by culturing the transformants under conditions in which the pressure is progressively increased, 10 thereby leading to amplification (at its chromosomal integration site) of both the selection gene and the linked DNA that encodes chaperone fragment. Amplification is the process by which genes in greater demand for the production of a protein critical for growth, together with closely associated genes which may encode a desired protein, are reiterated in tandem within the chromosomes of recombinant cells. Increased 15 quantities of desired protein are usually synthesised from thus amplified DNA.

Expression vectors usually contain a promoter that is recognised by the host organism and is operably linked to chaperone fragment nucleic acid. Such a promoter may be inducible or constitutive. The promoters are operably linked to DNA encoding 20 chaperone fragment by removing the promoter from the source DNA by restriction enzyme digestion and inserting the isolated promoter sequence into the vector. Both the native chaperone fragment promoter sequence and many heterologous promoters may be used to direct amplification and/or expression of chaperone fragment DNA. The term "operably linked" refers to a juxtaposition wherein the components described are 25 in a relationship permitting them to function in their intended manner. A control sequence "operably linked" to a coding sequence is ligated in such a way that expression of the coding sequence is achieved under conditions compatible with the control sequences.

30 Promoters suitable for use with prokaryotic hosts include, for example, the  $\beta$ -lactamase and lactose promoter systems, alkaline phosphatase, the tryptophan (trp) promoter

system and hybrid promoters such as the tac promoter. Their nucleotide sequences have been published, thereby enabling the skilled worker operably to ligate them to DNA encoding chaperone fragment, using linkers or adaptors to supply any required restriction sites. Promoters for use in bacterial systems will also generally contain a  
5 Shine-Delgarno sequence operably linked to the DNA encoding chaperone fragment.

Preferred expression vectors are bacterial expression vectors which comprise a promoter of a bacteriophage such as phage  $\lambda$  or T7 which is capable of functioning in the bacteria. In one of the most widely used expression systems, the nucleic acid  
10 encoding the fusion protein may be transcribed from the vector by T7 RNA polymerase (Studier et al, Methods in Enzymol. 185; 60-89, 1990). In the *E. coli* BL21(DE3) host strain, used in conjunction with pET vectors, the T7 RNA polymerase is produced from the  $\lambda$ -lysogen DE3 in the host bacterium, and its expression is under the control of the IPTG inducible lac UV5 promoter. This system has been employed successfully for  
15 over-production of many proteins. Alternatively the polymerase gene may be introduced on a lambda phage by infection with an int- phage such as the CE6 phage which is commercially available (Novagen, Madison, USA). other vectors include vectors containing the lambda PL promoter such as PLEX (Invitrogen, NL) , vectors containing the trc promoters such as pTrcHisXpress<sup>TM</sup> (Invitrogen) or pTrc99  
20 (Pharmacia Biotech, SE) , or vectors containing the tac promoter such as pKK223-3 (Pharmacia Biotech) or PMAL (new England Biolabs, MA, USA).

Suitable promoting sequences for use with yeast hosts may be regulated or constitutive and are preferably derived from a highly expressed yeast gene, especially a  
25 *Saccharomyces cerevisiae* gene. Thus, the promoter of the TRP1 gene, the ADH1 or ADH2 gene, the acid phosphatase (PHO5) gene, a promoter of the yeast mating pheromone genes coding for the  $\alpha$ - or  $\alpha$ -factor or a promoter derived from a gene encoding a glycolytic enzyme such as the promoter of the enolase, glyceraldehyde-3-phosphate dehydrogenase (GAP), 3-phospho glycerate kinase (PGK), hexokinase,  
30 pyruvate decarboxylase, phosphofructokinase, glucose-6-phosphate isomerase, 3-phosphoglycerate mutase, pyruvate kinase, triose phosphate isomerase, phosphoglucose

isomerase or glucokinase genes, the *S. cerevisiae* GAL 4 gene, the *S. pombe* nmt 1 gene or a promoter from the TATA binding protein (TBP) gene can be used. Furthermore, it is possible to use hybrid promoters comprising upstream activation sequences (UAS) of one yeast gene and downstream promoter elements including a functional TATA box of another yeast gene, for example a hybrid promoter including the UAS(s) of the yeast PH05 gene and downstream promoter elements including a functional TATA box of the yeast GAP gene (PH05-GAP hybrid promoter). A suitable constitutive PH05 promoter is e.g. a shortened acid phosphatase PH05 promoter devoid of the upstream regulatory elements (UAS) such as the PH05 (-173) promoter element starting at nucleotide -173 and ending at nucleotide -9 of the PH05 gene.

Chaperone fragment gene transcription from vectors in mammalian hosts may be controlled by promoters derived from the genomes of viruses such as polyoma virus, adenovirus, fowlpox virus, bovine papilloma virus, avian sarcoma virus, 15 cytomegalovirus (CMV), a retrovirus and Simian Virus 40 (SV40), from heterologous mammalian promoters such as the actin promoter or a very strong promoter, e.g. a ribosomal protein promoter, and from the promoter normally associated with chaperone fragment sequence, provided such promoters are compatible with the host cell systems.

20 Transcription of a DNA encoding chaperone fragment by higher eukaryotes may be increased by inserting an enhancer sequence into the vector. Enhancers are relatively orientation and position independent. Many enhancer sequences are known from mammalian genes (e.g. elastase and globin). However, typically one will employ an enhancer from a eukaryotic cell virus. Examples include the SV40 enhancer on the late 25 side of the replication origin (bp 100-270) and the CMV early promoter enhancer. The enhancer may be spliced into the vector at a position 5' or 3' to chaperone fragment DNA, but is preferably located at a site 5' from the promoter.

Advantageously, a eukaryotic expression vector encoding chaperone fragment may 30 comprise a locus control region (LCR). LCRs are capable of directing high-level integration site independent expression of transgenes integrated into host cell

chromatin, which is of importance especially where the chaperone fragment gene is to be expressed in the context of a permanently-transfected eukaryotic cell line in which chromosomal integration of the vector has occurred, in vectors designed for gene therapy applications or in transgenic animals.

5

Eukaryotic expression vectors may also contain sequences necessary for the termination of transcription and for stabilising the mRNA. Such sequences are commonly available from the 5' and 3' untranslated regions of eukaryotic or viral DNAs or cDNAs. These regions contain nucleotide segments transcribed as polyadenylated fragments in the 10 untranslated portion of the mRNA encoding chaperone fragment.

An expression vector includes any vector capable of expressing chaperone fragment nucleic acids that are operatively linked with regulatory sequences, such as promoter regions, that are capable of expression of such DNAs. Thus, an expression vector

15 refers to a recombinant DNA or RNA construct, such as a plasmid, a phage, recombinant virus or other vector, that upon introduction into an appropriate host cell, results in expression of the cloned DNA. Appropriate expression vectors are well known to those with ordinary skill in the art and include those that are replicable in eukaryotic and/or prokaryotic cells and those that remain episomal or those which 20 integrate into the host cell genome. For example, DNAs encoding chaperone fragment may be inserted into a vector suitable for expression of cDNAs in mammalian cells, e.g. a CMV enhancer-based vector such as pEVRF (Matthias, et al., (1989) NAR 17, 6418).

25 Particularly useful for practising the present invention are expression vectors that provide for the transient expression of DNA encoding chaperone fragment in mammalian cells. Transient expression usually involves the use of an expression vector that is able to replicate efficiently in a host cell, such that the host cell accumulates many copies of the expression vector, and, in turn, synthesises high levels of chaperone 30 fragment. For the purposes of the present invention, transient expression systems are

useful e.g. for identifying chaperone fragment mutants, to identify potential phosphorylation sites, or to characterise functional domains of the protein.

Construction of vectors according to the invention employs conventional ligation techniques. Isolated plasmids or DNA fragments are cleaved, tailored, and religated in the form desired to generate the plasmids required. If desired, analysis to confirm correct sequences in the constructed plasmids is performed in a known fashion. Suitable methods for constructing expression vectors, preparing *in vitro* transcripts, introducing DNA into host cells, and performing analyses for assessing chaperone fragment expression and function are known to those skilled in the art. Gene presence, amplification and/or expression may be measured in a sample directly, for example, by conventional Southern blotting, Northern blotting to quantitate the transcription of mRNA, dot blotting (DNA or RNA analysis), or *in situ* hybridisation, using an appropriately labelled probe which may be based on a sequence provided herein. Those skilled in the art will readily envisage how these methods may be modified, if desired.

The invention may be applied to the screening of molecular chaperone fragments for *in vivo* activity in a variety of ways. In general, a screening system requires a read-out of the activity of the agent being tested, and in the case of molecular chaperones having a general activity, cell viability or cell growth is affected by the chaperones and is easily monitored. In the case of more specific chaperones, however, it may be necessary to provide, in the cell, a reporter system whose activity is linked to the activity of the chaperone. For example, if the molecular chaperone in question facilitates the folding of a particular protein, a reporter system responsive to the protein in its correctly folded state may be used. For instance, where the protein is a transcription factor or can be incorporated into a chimeric transcription factor, a reporter system which is transactivatable by a transcription factor may be used.

Suitable reporter systems include those based on easily detectable gene products, such as luciferase, as  $\beta$  galactosidase or chloramphenicol acetyltransferase (CAT).

In the context of a screening system, the invention may be configured, for example, such that the transfected chaperone fragment is detected if it can rescue cell growth in a cell with a deficient endogenous chaperone activity. For example, growth of *E. coli* is dependent on GroEL activity. Temperature sensitive GroEL mutants will prevent *E. coli* from growing at non-permissive temperatures. Thus, *E. coli* may be transfected with vectors encoding chaperone fragments and assessed for growth at a non-permissive temperature, which is indicative of *in vivo* chaperone fragment activity.

Alternatively, a specialised chaperone required for the correct folding of a chosen polypeptide may be rendered deficient in the cell. In such a case, *in vivo* activity of a chaperone fragment may be assessed by measuring the biological activity of the chosen polypeptide, or by using a reporter system dependent on the chosen polypeptide.

Alternatively, where the chaperone has an observed activity such as an antiviral activity, for example in the case of GroEL which inhibits phage λ replication, a reporter system may be based on a viral promoter or replication origin.

The use of different reporter systems is likely to identify different *in vivo* activities of chaperone fragments. For example, in the case of GroEL, it is found that 193-335 complements mutants which have very low levels of endogenous GroEL, whilst 191-345 complements temperature sensitive mutants of GroEL. Fragment 191-376 is able to prevent phage λ replication. Thus, in a preferred aspect of the invention, the assay used to detect *in vivo* chaperone activity should reflect the use to which it is desired to put a minichaperone.

25

Endogenous chaperone proteins may be rendered deficient by a variety of means, including deletion of genes by homologous recombination, use of antisense molecules or specific inhibitors. The practice of these techniques is known in the art. Preferred, however, is the use of homologous recombination to disrupt endogenous genes.

30

In a preferred aspect, the invention may be applied not only to fragments of molecular chaperones, but also to mutants and variants of such fragments. The variants provided by the present invention include splice variants encoded by mRNA generated by alternative splicing of a primary transcript, amino acid mutants, glycosylation variants 5 and other covalent derivatives of the molecular chaperone which retain the physiological and/or physical properties thereof. Exemplary derivatives include molecules wherein the chaperone polypeptide is covalently modified by substitution, chemical, enzymatic, or other appropriate means with a moiety other than a naturally occurring amino acid. Such a moiety may be a detectable moiety such as an enzyme or 10 a radioisotope. Further included are naturally occurring variants of molecular chaperone found within a particular species, preferably a mammal. Such a variant may be encoded by a related gene of the same gene family, by an allelic variant of a particular gene, or represent an alternative splicing variant of a molecular chaperone gene.

15

Variants of the molecular chaperone fragments also comprise mutants thereof, which may contain amino acid deletions, additions or substitutions, subject to the requirement to maintain the activity of the molecular chaperone described herein. Thus, conservative amino acid substitutions may be made substantially without altering the 20 nature of the molecular chaperone, as may truncations from the 5' or 3' ends. Deletions and substitutions may moreover be made to the fragments of the molecular chaperone comprised by the invention. molecular chaperone mutants may be produced from a DNA encoding the molecular chaperone which has been subjected to in vitro mutagenesis resulting e.g. in an addition, exchange and/or deletion of one or more 25 amino acids. For example, substitutional, deletional or insertional variants of the molecular chaperone can be prepared by recombinant methods and screened for immuno-crossreactivity with the native forms of the molecular chaperone.

The fragments, mutants and other derivatives of the molecular chaperone preferably 30 retain substantial homology with the molecular chaperone . As used herein, "homology" means that the two entities share sufficient characteristics for the skilled

person to determine that they are similar in origin and function. Preferably, homology is used to refer to sequence identity.

"Substantial homology", where homology indicates sequence identity, means more than 5 40% sequence identity, preferably more than 45% sequence identity and most preferably a sequence identity of 50% or more, as judged by direct sequence alignment and comparison.

Sequence homology (or identity) may moreover be determined using any suitable 10 homology algorithm, using for example default parameters. Advantageously, the BLAST algorithm is employed, with parameters set to default values. The BLAST algorithm is described in detail at [http://www.ncbi.nih.gov/BLAST/blast\\_help.html](http://www.ncbi.nih.gov/BLAST/blast_help.html), which is incorporated herein by reference. The search parameters are defined as follows, and are advantageously set to the defined default parameters.

15

Advantageously, "substantial homology" when assessed by BLAST equates to sequences which match with an EXPECT value of at least about 7, preferably at least about 9 and most preferably 10 or more. The default threshold for EXPECT in BLAST searching is usually 10.

20

BLAST (Basic Local Alignment Search Tool) is the heuristic search algorithm employed by the programs blastp, blastn, blastx, tblastn, and tblastx; these programs ascribe significance to their findings using the statistical methods of Karlin and Altschul (see [http://www.ncbi.nih.gov/BLAST/blast\\_help.html](http://www.ncbi.nih.gov/BLAST/blast_help.html)) with a few enhancements. The 25 BLAST programs were tailored for sequence similarity searching, for example to identify homologues to a query sequence. The programs are not generally useful for motif-style searching. For a discussion of basic issues in similarity searching of sequence databases, see Altschul et al. (1994) *Nature Genetics* 6:119-129.

30 The five BLAST programs available at <http://www.ncbi.nlm.nih.gov> perform the following tasks:

**blastp** compares an amino acid query sequence against a protein sequence database;

**blastn** compares a nucleotide query sequence against a nucleotide sequence database;

5

**blastx** compares the six-frame conceptual translation products of a nucleotide query sequence (both strands) against a protein sequence database;

**tblastn** compares a protein query sequence against a nucleotide sequence database

10 dynamically translated in all six reading frames (both strands).

**tblastx** compares the six-frame translations of a nucleotide query sequence against the six-frame translations of a nucleotide sequence database.

15 **BLAST** uses the following search parameters:

**HISTOGRAM** Display a histogram of scores for each search; default is yes. (See parameter H in the BLAST Manual).

20 **DESCRIPTIONS** Restricts the number of short descriptions of matching sequences reported to the number specified; default limit is 100 descriptions. (See parameter V in the manual page). See also **EXPECT** and **CUTOFF**.

25 **ALIGNMENTS** Restricts database sequences to the number specified for which high-scoring segment pairs (HSPs) are reported; the default limit is 50. If more database sequences than this happen to satisfy the statistical significance threshold for reporting (see **EXPECT** and **CUTOFF** below), only the matches ascribed the greatest statistical significance are reported. (See parameter B in the BLAST Manual).

30 **EXPECT** The statistical significance threshold for reporting matches against database sequences; the default value is 10, such that 10 matches are expected to be found

merely by chance, according to the stochastic model of Karlin and Altschul (1990). If the statistical significance ascribed to a match is greater than the EXPECT threshold, the match will not be reported. Lower EXPECT thresholds are more stringent, leading to fewer chance matches being reported. Fractional values are acceptable. (See 5 parameter E in the BLAST Manual).

CUTOFF Cutoff score for reporting high-scoring segment pairs. The default value is calculated from the EXPECT value (see above). HSPs are reported for a database sequence only if the statistical significance ascribed to them is at least as high as would 10 be ascribed to a lone HSP having a score equal to the CUTOFF value. Higher CUTOFF values are more stringent, leading to fewer chance matches being reported. (See parameter S in the BLAST Manual). Typically, significance thresholds can be more intuitively managed using EXPECT.

15 MATRIX Specify an alternate scoring matrix for BLASTP, BLASTX, TBLASTN and TBLASTX. The default matrix is BLOSUM62 (Henikoff & Henikoff, 1992). The valid alternative choices include: PAM40, PAM120, PAM250 and IDENTITY. No alternate scoring matrices are available for BLASTN; specifying the MATRIX directive in BLASTN requests returns an error response.

20 STRAND Restrict a TBLASTN search to just the top or bottom strand of the database sequences; or restrict a BLASTN, BLASTX or TBLASTX search to just reading frames on the top or bottom strand of the query sequence.

25 FILTER Mask off segments of the query sequence that have low compositional complexity, as determined by the SEG program of Wootton & Federhen (1993) Computers and Chemistry 17:149-163, or segments consisting of short-periodicity internal repeats, as determined by the XNU program of Claverie & States (1993) Computers and Chemistry 17:191-201, or, for BLASTN, by the DUST program of 30 Tatusov and Lipman (see <http://www.ncbi.nlm.nih.gov>). Filtering can eliminate statistically significant but biologically uninteresting reports from the blast output (e.g.,

hits against common acidic-, basic- or proline-rich regions), leaving the more biologically interesting regions of the query sequence available for specific matching against database sequences.

- 5 Low complexity sequence found by a filter program is substituted using the letter "N" in nucleotide sequence (e.g., "NNNNNNNNNNNNNN") and the letter "X" in protein sequences (e.g., "XXXXXXXXXX").

Filtering is only applied to the query sequence (or its translation products), not to  
10 database sequences. Default filtering is DUST for BLASTN, SEG for other programs.

It is not unusual for nothing at all to be masked by SEG, XNU, or both, when applied to sequences in SWISS-PROT, so filtering should not be expected to always yield an effect. Furthermore, in some cases, sequences are masked in their entirety, indicating  
15 that the statistical significance of any matches reported against the unfiltered query sequence should be suspect.

NCBI-gi Causes NCBI gi identifiers to be shown in the output, in addition to the accession and/or locus name.

20 Most preferably, sequence comparisons are conducted using the simple BLAST search algorithm provided at <http://www.ncbi.nlm.nih.gov/BLAST>.

The various classes of chaperone proteins are generally homologous in structure and so  
25 there are therefore conserved or substantially homologous amino acid sequences between the members of the class. For instance, GroEL is just an example of an hsp60 chaperonin protein; other suitable proteins having an homologous apical domain may be followed.

The list was compiled from the OWL database release 28.1. The sequences listed below show clear homology to apical domain (residues 191-375) in PDB structure pdb1gr1.ent.

- 5 OWL is a non redundant database merging SWISS-PROT, PIR (1-3), GenBank (translation) and NRL-3D.

190-374 CH60\_ECOLI 60 KD CHAPERONIN (PROTEIN CPN60) (GROEL PROTEIN) (AMS). - ESCHERICHIA 190-374 CH60\_SALTI 60 KD  
10 CHAPERONIN (PROTEIN CPN60) (GROEL PROTEIN). - SALMONELLA TYPHI. 191-375 S56371 GroEL protein - Escherichia coli 190-374 CH60\_LEPIN 60 KD CHAPERONIN (PROTEIN CPN60) (GROEL PROTEIN) (HEAT SHOCK 58 KD PRO 191-375 S47530 groEL protein - Porphyromonas gingivalis 190-374 LPNHTPBG LPNHTPBG  
15 NID:g149691 - Legionella pneumophila (strain SVir) (library: 189-373 CH60\_ACTAC 60 KD CHAPERONIN (PROTEIN CPN60) (GROEL PROTEIN). - ACTINOBACILLUS ACT 191-375 JC4519 heat-shock protein GroEL - Pasteurella multocida  
191-375 CH60\_BRUAB 60 KD CHAPERONIN (PROTEIN CPN60) (GROEL PROTEIN). - BRUCELLA ABORTUS. 191-375 CH60\_HAEIN 60 KD  
20 CHAPERONIN (PROTEIN CPN60) (GROEL PROTEIN). - HAEMOPHILUS INFLUE 190-373 CH60\_CAUCR 60 KD CHAPERONIN (PROTEIN CPN60) (GROEL PROTEIN). - CAULOBACTER CRESCE 190-374 CH60\_AMOPS 60 KD CHAPERONIN (PROTEIN CPN60) (GROEL PROTEIN). -  
25 AMOEBA PROTEUS SYM 191-375 CH60\_HAEDU 60 KD CHAPERONIN (PROTEIN CPN60) (GROEL PROTEIN). - HAEMOPHILUS DUCREY 191-375 CH61\_RHIME 60 KD CHAPERONIN A (PROTEIN CPN60 A) (GROEL PROTEIN A). - RHIZOBIUM ME 190-374 CH60\_LEGMI 60 KD  
CHAPERONIN (PROTEIN CPN60) (GROEL PROTEIN) (58 KD COMMON  
30 ANTIGEN 191-375 CH60\_YEREN 60 KD CHAPERONIN (PROTEIN CPN60) (GROEL PROTEIN) (HEAT SHOCK PROTEIN 6) 190-374 CH

63\_BRAJA 60 KD CHAPERONIN 3 (PROTEIN CPN60 3) (GROEL PROTEIN 3). - BRADYRHIZOBI 191-375 CH60\_PORGI 60 KD CHAPERONIN (PROTEIN CPN60) (GROEL PROTEIN). - PORPHYROMONAS GING 191-375 S52901 heat shock protein 60K - Yersinia enterocolitica  
5 191-375 S26423 heat shock protein 60 - Yersinia enterocolitica  
191-375 RSU373691 RSU37369 NID: g1208541 - Rhodobacter sphaeroides strain=HR. 190-374 CH62\_BRAJA 60 KD CHAPERONIN 2 (PROTEIN CPN60 2) (GROEL PROTEIN 2). - BRADYRHIZOBI 191-375  
10 CH60\_ACYPS 60 KD CHAPERONIN (PROTEIN CPN60) (GROEL PROTEIN) (SYMBIONIN). - ACYRTH 191-375 CH63\_RHIME 60 KD CHAPERONIN C (PROTEIN CPN60 C) (GROEL PROTEIN C). - RHIZOBIUM ME 191-375 YEPHSPCRP1 YEPHSPCRP NID: g466575 - Yersinia enterocolitica DNA. 191-375 CH60\_BORPE 60 KD CHAPERONIN  
15 (PROTEIN CPN60) (GROEL PROTEIN). - BORDETELLA PERTUSS 189-373 BRUGRO1 BRUGRO NID: g144106 - Brucella aabortus (library: lambda-2001) DNA.  
191-375 CH60\_PSEAE 60 KD CHAPERONIN (PROTEIN CPN60) (GROEL PROTEIN). - PSEUDOMONAS AERUGI 190-374 CH60\_BARBA 60 KD  
20 CHAPERONIN (PROTEIN CPN60) (IMMUNOREACTIVE PROTEIN BB65) (IMMUNO 191-375 BAOBB63A BAOBB63A NID: g143845 - Bartonella bacilliformis (library: ATCC 35685) 189-373 CH60\_BACST 60 KD CHAPERONIN (PROTEIN CPN60) (GROEL PROTEIN). - BACILLUS STEAROTHE 188-372  
25 190-373 CH60\_BORBU 60 KD CHAPERONIN (PROTEIN CPN60) (GROEL PROTEIN). - BORRELIA BURGDORFE 224-408 S26583 chaperonin hsp60 - maize 190-373 A49209 heat shock protein HSP60 - Lyme disease spirochete 224-408 MZECPN60B MZECPN60B NID: g309558 - Zea mays (strain B73) (library: Dash11 of P.S 189-373  
30 CH60\_THEP3 60 KD CHAPERONIN (PROTEIN CPN60) (GROEL PROTEIN) (HEAT SHOCK 61 KD PRO 188-372 CH60\_STAEP 60 KD

CHAPERONIN (PROTEIN CPN60) (GROEL PROTEIN) (HEAT SHOCK PROTEIN  
6 189-373 CH60\_LACLA 60 KD CHAPERONIN (PROTEIN CPN60) (GROEL  
PROTEIN). - LACTOCOCCUS LACTIS 188-374 CH61\_STRAL 60 KD  
CHAPERONIN 1 (PROTEIN CPN60 1) (GROEL PROTEIN 1) (HSP58). -  
5 STRE 191-375 CH60\_CHLPN 60 KD CHAPERONIN (PROTEIN  
CPN60) (GROEL PROTEIN). - CHLAMYDIA PNEUMONI 224-408  
MZECPN60A MZECPN60A NID: g309556 - Zea mays (strain  
B73) (library:Dach 11 of P. 190-373 HECHSPAB1 HECHSPAB NID:  
g712829 - Helicobacter pylori (individual\_isolate 85P) D  
10 221-405 CH60\_ARATH MITOCHONDRIAL CHAPERONIN HSP60 PRECURSOR.  
- ARABIDOPSIS THALIANA (MOUS 224-408 CH60\_MAIZE  
MITOCHONDRIAL CHAPERONIN HSP60 PRECURSOR. - ZEA MAYS  
(MAIZE). 190-374 CH60\_CHLTR 60 KD CHAPERONIN (PROTEIN  
CPN60) (GROEL PROTEIN) (57 KD CHLAMYDIAL HYP 189-373  
15 CH60\_STAAU 60 KD CHAPERONIN (PROTEIN CPN60) (GROEL  
PROTEIN) (HEAT SHOCK PROTEIN 6 189-373 CH60\_CLOPE 60 KD  
CHAPERONIN (PROTEIN CPN60) (GROEL PROTEIN). - CLOSTRIDIUM  
PERFRI 212-397 HS60\_YEAST HEAT SHOCK PROTEIN 60 PRECURSOR  
(STIMULATOR FACTOR 1 66 KD COMPONENT) 217-403 CH60\_PYRSA 60  
20 KD CHAPERONIN (PROTEIN CPN60) (GROEL PROTEIN). - PYRENOMONAS  
SALINA 191-377 CH60\_EHRCH 60 KD CHAPERONIN (PROTEIN  
CPN60) (GROEL PROTEIN). - EHRLICHIA CHAFFEEN 191-375 CHTGROE1  
CHTGROE NID: g144503 - C.trachomatis DNA. 188-372 CH60\_THETH  
60 KD CHAPERONIN (PROTEIN CPN60) (GROEL PROTEIN). - THERMUS  
25 AQUATICUS 189-373 TAU294831 TAU29483 NID: g1122940 - Thermus  
aquaticus. 190-378 CH60\_RICTS 60 KD CHAPERONIN (PROTEIN  
CPN60) (GROEL PROTEIN) (MAJOR ANTIGEN 58) (5 189-375 SYCCPNC  
SYCCPNC NID: g1001102 - Synechocystis sp. (strain PCC6803,)  
DNA.  
30 190-373 CPU308211 CPU30821 NID: g1016083 - Cyanophora  
paradoxa. 189-373 CH61\_MYCLE 60 KD CHAPERONIN 1 (PROTEIN

CPN60 1) (GROEL PROTEIN 1). - MYCOBACTERIUM 239-423 PSU21139  
PSU21139 NID: g806807 - pea. 191-377 CH60\_COWRU 60 KD  
CHAPERONIN (PROTEIN CPN60) (GROEL PROTEIN). - COWDRIA  
RUMINANTIUM 245-429 RUBB\_BRANA RUBISCO SUBUNIT BINDING-  
5 PROTEIN BETA SUBUNIT PRECURSOR (60 KD CHAPERON 144-328  
SCCPN60 SCCPN60 NID: g1167857 - rye.  
153-338 CH60\_EHRR1 60 KD CHAPERONIN (PROTEIN CPN60) (GROEL  
PROTEIN) (55 KD MAJOR ANTIGEN) 245-429 RUBB\_ARATH RUBISCO  
SUBUNIT BINDING-PROTEIN BETA SUBUNIT PRECURSOR (60 KD  
10 CHAPERON 235-419 ATU49357 ATU49357 NID: g1223909 - thale  
cress strain=ecotype Wassilewskija. 195-379 RUB1\_BRANA  
RUBISCO SUBUNIT BINDING-PROTEIN ALPHA SUBUNIT (60 KD  
CHAPERONIN ALPHA 189-374 CH62\_SYN1 60 KD CHAPERONIN 2  
(PROTEIN CPN60 2) (GROEL HOMOLOG 2). - SYNECHOCYSTI 178-362  
15 RUBA\_RICCO RUBISCO SUBUNIT BINDING-PROTEIN ALPHA SUBUNIT (60  
KD CHAPERONIN ALPHA 190-375 CH60\_ODOSI 60 KD CHAPERONIN  
(PROTEIN CPN60) (GROEL PROTEIN). - ODONTELLA SINENSIS 236-420  
PSU21105 PSU21105 NID: g1185389 - pea. 224-409 CH60\_BRANA  
MITOCHONDRIAL CHAPERONIN CH60\_BACSU 60 KD CHAPERONIN  
20 (PROTEIN CPN60) (GROEL PROTEIN). - BACILLUS SUBTILIS. 191-375  
CH60\_AGRTU 60 KD CHAPERONIN (PROTEIN CPN60) (GROEL PROTEIN).  
- AGROBACTERIUM TUME 191-375 b36917 heat shock protein GroEL  
- Agrobacterium tumefaciens  
191-375 PAU17072 PAU17072 NID: g576778 - Pseudomonas  
25 aeruginosa. 191-375 CH60\_RHILV 60 KD CHAPERONIN (PROTEIN  
CPN60) (GROEL PROTEIN). - RHIZOBIUM LEGUMINO 187-373  
CH61\_STRCO 60 KD CHAPERONIN 1 (PROTEIN CPN60 1) (GROEL  
PROTEIN 1) (HSP58). - STRE 191-375 CH60\_COXBU 60 KD CHAPERONIN  
(PROTEIN CPN60) (GROEL PROTEIN) (HEAT SHOCK PROTEIN B 191-375  
30 CH62\_RHIME 60 KD CHAPERONIN B (PROTEIN CPN60 B) (GROEL  
PROTEIN B). - RHIZOBIUM ME 191-375 PSEGROESL1 PSEGROESL NID:

g151241 - *Pseudomonas aeruginosa* (library: ATCC 27853) 189-  
372 CH61\_SYNY3 60 KD CHAPERONIN 1 (PROTEIN CPN60 1) (GROEL  
HOMOLOG 1). - *SYNECHOCYSTI* 189-373 CH60\_CLOTM 60 KD CHAPERONIN  
(PROTEIN CPN60) (GROEL PROTEIN) (HSP-60). - CLOSTRIDI 191-373  
5 CH60\_PSEPU 60 KD CHAPERONIN (PROTEIN CPN60) (GROEL PROTEIN). -  
PSEUDOMONAS PUTIDA 190-373 CH60\_SYNP7 60 KD CHAPERONIN  
(PROTEIN CPN60) (GROEL PROTEIN). - *SYNECHOCOCCUS* SP. 190-374  
CH60\_GALSU 60 KD CHAPERONIN (PROTEIN CPN60) (GROEL PROTEIN). -  
GALDIERIA SULPHURA 190-374 CH60\_ZYMMO 60 KD CHAPERONIN  
10 (PROTEIN CPN60) (GROEL PROTEIN). - *ZYMOMONAS MOBILIS*. 191-375  
JC2564 heat shock protein groEL - *Zymomonas mobilis*  
191-375 CH60\_CHRVI 60 KD CHAPERONIN (PROTEIN CPN60) (GROEL  
PROTEIN). - *CHROMATIUM VINOsum* 189-373 CH60\_MYCTU 60 KD  
CHAPERONIN (PROTEIN CPN60) (GROEL PROTEIN) (65 KD  
15 ANTIGEN) (HEAT 191-375 CH60\_NEIME 60 KD CHAPERONIN (PROTEIN  
CPN60) (GROEL PROTEIN) (63 KD STRESS PROTEIN 189-373  
CH60\_TREPA 60 KD CHAPERONIN (PROTEIN CPN60) (GROEL  
PROTEIN) (TPN60) (TP4 ANTIGEN) 190-374 CH60\_HELPY 60 KD  
CHAPERONIN (PROTEIN CPN60) (GROEL PROTEIN) (HEAT SHOCK PROTEIN  
20 6 191-375 CH60\_NEIGO 60 KD CHAPERONIN (PROTEIN CPN60) (GROEL  
PROTEIN) (63 KD STRESS PROTEIN 222-406 CH61\_CUCMA  
MITOCHONDRIAL CHAPERONIN HSP60-1 PRECURSOR. - *CUCURBITA*  
MAXIMA (PUMPKI 189-373 CH60\_MYCPA 60 KD CHAPERONIN (PROTEIN  
CPN60) (GROEL PROTEIN) (65 KD ANTIGEN) (HEAT 230-414 MPU15989  
25 MPU15989 NID:g559802 - *Mycobacterium paratuberculosis*. 224-  
408 S26582 chaperonin hsp60 - maize 191-375 S40247 heat-  
shock protein - *Neisseria gonorrhoeae* 189-373 CH60\_CLOAB 60  
KD CHAPERONIN (PROTEIN CPN60) (GROEL PROTEIN). - CLOSTRIDIUM  
ACETOB 191-375 CH60\_NEIFL 60 KD CHAPERONIN (PROTEIN  
30 CPN60) (GROEL PROTEIN) (63 KD STRESS PROTEIN 190-373  
CH60\_LEGPN 60 KD CHAPERONIN (PROTEIN CPN60) (GROEL

PROTEIN) (58 KD COMMON ANTIGEN 222-406 CH62\_CUCMA  
MITOCHONDRIAL CHAPERONIN HSP60-2 PRECURSOR. - CUCURBITA  
MAXIMA (PUMPKI 191-375 CHTGROESL1 CHTGROESL NID: g402332 -  
Chlamydia trachomatis DNA. 64-248 S40172 S40172 NID:  
5 g251679 - Chlamydia psittaci pigeon strain P-1041. 189-373  
SYOGROEL2 SYOGROEL2 NID:g562270 - Synechococcus vulcanus  
DNA. 191-375 CH60\_CHLPS 60 KD CHAPERONIN (PROTEIN  
CPN60) (GROEL PROTEIN) (57 KD CHLAMYDIAL HYP 188-372  
CH62\_STRAL 60 KD CHAPERONIN 2 (PROTEIN CPN60 2) (GROEL  
10 PROTEIN 2) (HSP56). - STRE 189-373 CH62\_MYCLE 60 KD  
CHAPERONIN 2 (PROTEIN CPN60 2) (GROEL PROTEIN 2) (65 KD  
ANTIGEN) 236-420 MSGANTM MSGANTM NID: g149923 - M.leprae  
DNA, clone Y3178.  
CPN60 PRECURSOR. - BRASSICA NAPUS (RAPE). 105-289 PMSARG2  
15 PMSARG2 NID: g607157 - Prochlorococcus marinus.  
234-417 RUB2\_BRANA RUBISCO SUBUNIT BINDING-PROTEIN ALPHA  
SUBUNIT PRECURSOR (60 KD CHAPERO 75-259 CRECPN1A CRECPN1A  
NID: g603910 - Chlamydomonas reinhardtii cDNA to mRNA. 215-  
400 P60\_CRIGR MITOCHONIDRIAL MATRIX PROTEIN P1 PRECURSOR  
20 (P60 LYMPHOCYTE PROTEIN) (CH224-408 CRECPN1B CRECPN1B NID:  
g603912 - Chlamydomonas reinhardtii cDNA to mRNA. 191-375  
RUBA\_WHEAT RUBISCO SUBUNIT BINDING-PROTEIN ALPHA SUBUNIT  
PRECURSOR (60 KD CHAPERO 189-373 B47292 heat shock protein  
groEL - Mycobacterium tuberculosis  
25 206-391 CELHSP60CP CELHSP60CP NID: g533166 - Caenorhabditis  
elegans (strain CB1392) cDNA 215-400 P60\_HUMAN MITOCHONDRIAL  
MATRIX PROTEIN P1 PRECURSOR (P60 LYMPHOCYTE PROTEIN) (CH 215-  
400 P60\_MOUSE MITOCHONDRIAL MATRIX PROTEIN P1 PRECURSOR (P60  
LYMPHOCYTE PROTEIN) (CH 215-400 P60\_RAT MITOCHONDRIAL MATRIX  
30 PROTEIN P1 PRECURSOR (P60 LYMPHOCYTE PROTEIN) (CH 215-400  
A41931 chaperonin hsp60 - mouse

30

- 197-382 MMHSP60A MMHSP60A NID:g51451 - house mouse. 218-402  
CH63\_HELVI 63 KD CHAPERONIN PRECURSOR (P63). - HELIOTHIS  
VIRESSENS (NOCTUID MOTH) 205-390 EGHSP60GN EGHSP60GN NID:  
g1217625 - Euglena gracilis. 222-407 HS60\_SCHPO PROBABLE  
5 HEAT SHOCK PROTEIN 60 PRECURSOR. - SCHIZOSACCHAROMYCES POMBE  
198-385 S61295 heat shock protein 60 - Trypanosoma cruzi  
198-385 TRBMTHSP TRBMTHSP NID: g903883 - Mitochondrion  
Trypanosoma brucei (strain EATRO 8-69 ECOGROELA ECOGROELA  
NID: g146268 - E.coli DNA, clone E. 142-325 ENHCPN60P  
10 ENHCPN60P NID: g675513 - Entamoeba histolytica (strain HM-  
1:IMSS) DNA. 257-433 CH60\_PLAFG MITOCHONDRIAL CHAPERONIN  
CPN60 PRECURSOR. - PLASMODIUM FALCIPARUM (ISO 1-90 CRECPN1C  
CRECPN1C NID: g603914 - Chlamydomonas reinhardtii cDNA to  
mRNA.  
15 5-65 ATTS0779 ATTS0779 NID: g17503 - thale cress.  
189-373 CH60\_MYCGE 60 KD CHAPERONIN (PROTEIN CPN60) (GROEL  
PROTEIN). - MYCOPLASMA GENITAL 228-411 HTOHSP60X HTOHSP60X  
NID: g553068 - Histoplasma capsulatum (strain G217B) DNA.  
190-297 CH60\_SYNPG6 60 KD CHAPERONIN (PROTEIN CPN60) (GROEL  
20 PROTEIN) (FRAGMENT). - SYNECHO 169-245 RUBA\_ARATH RUBISCO  
SUBUNIT BINDING-PROTEIN ALPHA SUBUNIT (60 KD CHAPERONIN  
ALPHA.)

Such analyses may be repeated using other databases, or more recent updates of the  
25 OWL database, and for other chaperone families, such as the HSP 70, HSP 90 or GRP  
families.

In a further aspect, the invention provides a method for providing a chaperone activity  
*in vivo*, comprising administering to a cell a fragment of a molecular chaperone which  
30 has *in vivo* activity.

Fragments of molecular chaperones which possess *in vivo* activity may be identified by the method of the first aspect of the invention. Preferred fragments are GroEL fragments 191-345 and 193-335.

- 5    Fragments according to the invention are capable of complementing deficiencies in the endogenous molecular chaperones of cells and are therefore particularly suited for rectifying defects caused by chaperone deficiencies. Preferably, fragments of molecular chaperones having *in vivo* activity may be administered to patients suffering from disorders or diseases associated with anomalies in protein folding or other chaperone  
10    functions.

The active ingredients of a pharmaceutical composition comprising the chaperone fragment are contemplated to exhibit excellent therapeutic activity, for example, in the alleviation of Alzheimer's disease when administered in amount which depends on the  
15    particular case. Dosage regima may be adjusted to provide the optimum therapeutic response. For example, several divided doses may be administered daily or the dose may be proportionally reduced as indicated by the exigencies of the therapeutic situation.

- 20    The active compound may be administered in a convenient manner such as by the oral, intravenous (where water soluble), intramuscular, subcutaneous, intranasal, intradermal or suppository routes or implanting (e.g. using slow release molecules). Depending on the route of administration, the active ingredient may be required to be coated in a material to protect said ingredients from the action of enzymes, acids and other natural  
25    conditions which may inactivate said ingredient.

In order to administer the polypeptide by other than parenteral administration, it will be coated by, or administered with, a material to prevent its inactivation. For example, the polypeptide may be administered in an adjuvant, co-administered with enzyme  
30    inhibitors or in liposomes. Adjuvant is used in its broadest sense and includes any immune stimulating compound such as interferon. Adjuvants contemplated herein

include resorcinols, non-ionic surfactants such as polyoxyethylene oleyl ether and n-hexadecyl polyethylene ether. Enzyme inhibitors include pancreatic trypsin.

Liposomes include water-in-oil-in-water CGF emulsions as well as conventional  
5 liposomes.

The active compound may also be administered parenterally or intraperitoneally. Dispersions can also be prepared in glycerol, liquid polyethylene glycols, and mixtures thereof and in oils. Under ordinary conditions of storage and use, these preparations  
10 contain a preservative to prevent the growth of microorganisms.

The pharmaceutical forms suitable for injectable use include sterile aqueous solutions (where water soluble) or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersion. In all cases the form must be  
15 sterile and must be fluid to the extent that easy syringability exists. It must be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (for example, glycerol, propylene glycol, and liquid polyethyleneglycol, and the like),  
20 suitable mixtures thereof, and vegetable oils. The proper fluidity can be maintained, for example, by the use of a coating such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the use of surfactants.

The prevention of the action of microorganisms can be brought about by various  
25 antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, sorbic acid, thimerosal, and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars or sodium chloride. Prolonged absorption of the injectable compositions can be brought about by the use in the compositions of agents delaying absorption, for example, aluminium monostearate and gelatin.

- Sterile injectable solutions are prepared by incorporating the active compound in the required amount in the appropriate solvent with various of the other ingredients enumerated above, as required, followed by filtered sterilisation. Generally, dispersions are prepared by incorporating the sterilised active ingredient into a sterile vehicle which contains the basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, the preferred methods of preparation are vacuum drying and the freeze-drying technique which yield a powder of the active ingredient plus any additional desired ingredient from previously sterile-filtered solution thereof.
- When the chaperone fragment is suitably protected as described above, it may be orally administered, for example, with an inert diluent or with an assimilable edible carrier, or it may be enclosed in hard or soft shell gelatin capsules, or it may be compressed into tablets, or it may be incorporated directly with the food of the diet. For oral therapeutic administration, the active compound may be incorporated with excipients and used in the form of ingestible tablets, buccal tablets, troches, capsules, elixirs, suspensions, syrups, wafers, and the like. The amount of active compound in such therapeutically useful compositions in such that a suitable dosage will be obtained.
- The tablets, troches, pills, capsules and the like may also contain the following: a binder such as gum tragacanth, acacia, corn starch or gelatin; excipients such as dicalcium phosphate; a disintegrating agent such as corn starch, potato starch, alginic acid and the like; a lubricant such as magnesium stearate; and a sweetening agent such as sucrose, lactose or saccharin may be added or a flavouring agent such as peppermint, oil of wintergreen, or cherry flavouring. When the dosage unit form is a capsule, it may contain, in addition to materials of the above type, a liquid carrier.

Various other materials may be present as coatings or to otherwise modify the physical form of the dosage unit. For instance, tablets, pills, or capsules may be coated with shellac, sugar or both. A syrup or elixir may contain the active compound, sucrose as a sweetening agent, methyl and propylparabens as preservatives, a dye and flavouring

such as cherry or orange flavour. Of course, any material used in preparing any dosage unit form should be pharmaceutically pure and substantially non-toxic in the amounts employed. In addition, the active compound may be incorporated into sustained-release preparations and formulations.

5

As used herein "pharmaceutically acceptable carrier and/or diluent" includes any and all solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents and the like. The use of such media and agents for pharmaceutical active substances is well known in the art. Except insofar as any conventional media or agent is incompatible with the active ingredient, use thereof in the therapeutic compositions is contemplated. Supplementary active ingredients can also be incorporated into the compositions.

It is especially advantageous to formulate parenteral compositions in dosage unit form for ease of administration and uniformity of dosage. Dosage unit form as used herein refers to physically discrete units suited as unitary dosages for the mammalian subjects to be treated; each unit containing a predetermined quantity of active material calculated to produce the desired therapeutic effect in association with the required pharmaceutical carrier. The specification for the novel dosage unit forms of the invention are dictated by and directly dependent on (a) the unique characteristics of the active material and the particular therapeutic effect to be achieved, and (b) the limitations inherent in the art of compounding such as active material for the treatment of disease in living subjects having a diseased condition in which bodily health is impaired.

25 The principal active ingredient is compounded for convenient and effective administration in effective amounts with a suitable pharmaceutically acceptable carrier in dosage unit form. In the case of compositions containing supplementary active ingredients, the dosages are determined by reference to the usual dose and manner of administration of the said ingredients.

30

In a preferred aspect there is provided polypeptide of the invention as hereinbefore defined for use in the treatment of disease. Consequently there is provided the use of a chaperone fragment of the invention for the manufacture of a medicament for the treatment of disease associated with aberrant protein/polypeptide structure. The 5 aberrant nature of the protein/polypeptide may be due to misfolding or unfolding which in turn may be due to an anomalous e.g. mutated amino acid sequence. The protein/polypeptide may be destabilised or deposited as plaques e.g. as in Alzheimer's disease. The disease might be caused by a prion. A polypeptide-based medicament of the invention would act to renature or resolubilise aberrant, defective or deposited 10 proteins.

In a further aspect, there is provided a nucleic acid molecule encoding a chaperone fragment in accordance with other aspects of the invention for use in the treatment of disease. Consequently, there is provided the use of a nucleic acid molecule of the 15 invention for the manufacture of a medicament for the treatment of disease associated with protein/polypeptide structure. Genetic therapy *in vivo* is therefore provided for by way of introduction and expression of DNA encoding the chaperone fragment in cells/tissues of an individual to provide chaperonin activity in those cells/tissues.

20 The invention is further described, for the purposes of illustration only, in the following examples.

## Materials and Methods

25 **Bacterial and bacteriophage strains.** The *E. coli* strains used in this study are: C41(DE3), a mutant of BL21(DE3) capable of expressing toxic genes (Miroux, B. & Walker, J. E. (1996) *J. Mol. Biol.* **260**, 289-298); SV2 (B178<sup>groEL44</sup>) and SV6 (B178<sup>groEL673</sup>): isogenic strains carrying temperature-sensitive alleles of *groEL*; SV1(=B178) (24), AI90 ( $\Delta$ *groEL*::*kanR*) [pBAD-EL] (Ivic, A., Olden, D., Wallington, 30 E. J. & Lund, P. A. (1997) *Gene* **194**, 1-8), and TG1 (Gibson, T. J. (1984) Ph.D. thesis, University of Cambridge).  $\lambda$  b2cl (Zeilstra-Ryalls, J., Fayet, O., Baird, L. &

Georgopoulos, C. (1993) *J. Bacteriol.* **175**, 1134-1143) is used, according to standard methods (Arber, W., Enquist, L., Hohn, B., Murray, N. E. & Murray, K. (1983) in *Lambda II*, eds. Hendrix, R., Roberts, J., Stahl, F. & Weisberg, R. (Cold Spring Harbor Laboratory, N.Y.), pp. 433-466). Plaque formation is assayed at 30°C.

5

**Plasmid construction.** Standard molecular biology procedures are used (Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) *Molecular Cloning. A Laboratory Manual* (Cold Spring Harbor Laboratory, N.Y.)) . The principal steps of plasmid construction are summarised in Fig. 1.

10

Genomic DNA is extracted from *E. coli* TG1 (Gibson, T. J. (1984) Ph.D. thesis, University of Cambridge) using a commercial kit (Qiagen). PCR using two oligonucleotides 5' - ATT CAT ATG AAT ATT CGT CCA TTG CAT GAT CG - 3' (SEQ. ID. No. 1) and 5' - AA CGG CCG TTA ATT AAG **GTG** CAC CGA AAG ATT 15 TAT CCA GAA CTA CG - 3' (SEQ. ID. No. 2) produce a 494 bp DNA carrying the complete 294 bp of groES gene and unique sites for *Nde* I and *Eag* I (underlined). This fragment also contains the 44 bp intergenic ES-EL region and the first 123 bp of *groEL*, including the unique *Apal* I site (bold characters) of the groE operon (Hemmingsen, S. M., Woolford, C., van der Vies, S., Tilly, K., Dennis, D. T., 20 Georgopoulos, C. P., Hendrix, R. W. & Ellis, R. J. (1988) *Nature* **333**, 330-334). The *GroEL* gene (SEQ. ID. No. 10) is PCR amplified using two oligonucleotides 5' - T AGC TGC CAT ATG GCA GCT AAA GAC GTA AAA TTC GG - 3' (SEQ. ID. No. 3) and 5' - ATG TAA CGG CCG TTA CAT CAT GCC GCC CAT GCC ACC - 3' (SEQ. ID. No. 4) producing a 1,659 bp DNA with unique sites for *Nde* I and *Eag* I 25 (underlined). These fragments are cloned into pRSETA-*Eag* I allowing expression of GroE proteins under control of a T7 promoter. The unique *Eag*I recognition site (underlined) is created by replacing the *EcoR* I-*Hind* III fragment of pRSETA (InVitrogen) using a synthetic DNA cassette consisting of two oligos 5' - AAT TCA A CGG CCG TTA - 3' (SEQ. ID. No. 5) and 5' - AGC TTA ACG GCC GTT G- 3' (SEQ. 30 ID. No. 6). The pRSETA-*Eag* I (GroESL) vector is generated by subcloning the *Nde* I/*Apal* I fragment from pRSETA-*Eag* I (GroES) into pRSETA-*Eag* I (GroEL)  $\Delta$ *Nde* I-

*ApaL* I. Various N- and C-terminally truncated fragments of the apical domain of GroEL are cloned by PCR into *BamH* I and *Eag* I sites of pRSETA-*Eag* I vector encoding an N-terminal histidine-tail (17 amino acids; "sht"), MRGSHHHHHGLVPRGS (SEQ. ID. No. 7), which contains an engineered thrombin cleavage site (Zahn, R., Buckle, A. M., Perret, S., Johnson, C. M. J., Corrales, F. J., Golbik, R. & Fersht, A. R. (1996) *Proc. Natl. Acad. Sci. USA* 93, 15024-15029). The mutation Y203E is introduced into GroEL by PCR, as described (Hemsley, A., Arnheim, N., Toney, M. D., Cortopassi, G. & Galas, D. J. (1989) *Nucleic Acids Res.* 17, 6545-6551), using oligonucleotides 5' - ATT CAT CAA CAA GCC - 3' (SEQ. ID. No. 8) and 5' - TCA GGA GAC AGG TAG CC - 3' (SEQ. ID. No. 9), creating a unique *EcoR* I site (bold characters).

A modified pACYC184 vector (New England Biolabs) is constructed. The different pRSETA-*Eag* I based vectors are digested by *Xba* I, the recessed 3' ends filled in with Klenow enzyme and then, digested by *Eag* I. The *Xba* I blunt-ended-*Eag* I fragments, containing the ribosome binding site of pRSETA, are ligated into pACYC184 *EcoR* V / *Eag* I digested and alkaline phosphatase treated plasmid. pJC vector is generated by replacing the *Xmn* I/ *Ase* I fragment of pACYC184 by the *Ase* I-*BsaB* I of pBR322 prepared in the *dam*<sup>-</sup> *dcm*<sup>-</sup> JM110 *E. coli* strain (Yanisch-Perron, C., Viera, J. & Messing, J. (1985) *Gene* 33, 103-119). The *Xba* I / *Hind* III fragments from pRSETA-*Eag* I based vectors are cloned into pBAD30 *Xba* I / *Hind* III digested and alkaline phosphatase treated plasmid. Alternatively, the different groE genes are subcloned into the unique *Nde* I and *Eag* I unique sites of pACYC184 or pJC and pBAD30 (Guzman, L.-M., Belin, D., Carson, M. J. & Beckwith, J. (1995) *J. Bacteriol.* 177, 4121-4130) vectors. Plasmids containing no groE inserts are generated by ligated filled-in recessed 3' ends of pACYC184 or pJC and pBAD30  $\Delta$ *Nde* I-*Eag* I vectors.

A colony-based PCR procedure (Güssow, D. & Clackson, T. (1989) *Nucleic Acids Res.* 17: 4000) is performed to identify the positive clones using T7 promoter and 3' reverse cloning oligonucleotides. PCR cycle sequencing using fluorescent dideoxy chain

terminators (Applied Biosystems) are performed. Sequencing reactions are analysed on an Applied Biosystems 373A Automated DNA. All PCR amplified DNA fragments are sequenced after cloning. The *groEL* gene differs from the sequence in the database (Hemmingsen, S. M., Woolford, C., van der Vies, S., Tilly, K., Dennis, D. T., 5 Georgopoulos, C. P., Hendrix, R. W. & Ellis, R. J. (1988) *Nature* **333**, 330-334) by the substitutions A262L and I267M as described (Zeilstra-Ryalls, J., Fayet, O., Baird, L. & Georgopoulos, C. (1993) *J. Bacteriol.* **175**, 1134-1143).

**Protein production and characterisation.** GroE proteins, ~57.5 kDa GroEL and 10 ~10 kDa GroES, are expressed by inducing the T7 promoter of pRSETA-Eag I based vectors with isopropyl- $\beta$ -D-thiogalactoside (IPTG) in *E. coli* C41(DE3) (Miroux, B. & Walker, J. E. (1996) *J. Mol. Biol.* **260**, 289-298). Purification is performed as previously described (Corrales, F. J. & Fersht, A. R. (1996) *Folding & Design* **1**, 265-15 273). The over-expression and purification of minichaperones in *E. coli* C41(DE3) cells is carried out essentially as previously described (Zahn, R., Buckle, A. M., Perret, S., Johnson, C. M. J., Corrales, F. J., Golbik, R. & Fersht, A. R. (1996) *Proc. Natl. Acad. Sci. USA* **93**, 15024-15029). Proteins are analysed by electrospray mass spectrometry. Protein concentration is determined by absorbance at 276 nm using the method of Gill & von Hippel (Gill, S. C. & von Hippel, P. H. (1989) *Analyt. Biochem.* **182**, 319-326) and confirmed by quantitative amino acid analysis. Constitutive expression under the control of the tetracycline-resistance gene promoter / operator is obtained either using the low copy-number pACYC184 or the high copy-number pJC vectors. pBAD30 vector allows inducible expression with 0.2-0.5% arabinose controlled by the *PBAD* promoter and its regulatory gene, *araC* (Guzman, 20 L.-M., Belin, D., Carson, M. J. & Beckwith, J. (1995) *J. Bacteriol.* **177**, 4121-4130). The level of expression of GroEL minichaperones is analysed by 15% sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) under non-reducing conditions followed by Western blotting. After separation of proteins, GroEL molecules are detected with rabbit anti-GroEL antibodies (Sigma) followed by anti-25 rabbit immunoglobulins horseradish peroxidase conjugate antibodies (Sigma). 30

Bromochloroindolyl phosphate (BCIP) / nitro blue tetrazolium (NBT) chromogens (Sigma) are used as substrate.

5        **In vitro refolding experiments.** Refolding assays of rhodanese (Horowitz, P. M. (1995) in *Chaperonin-assisted protein folding of the enzyme rhodanese by GroEL/GroES*, eds. Shirley, B. A. (Humana Press), Vol. 40, pp. 361-368) and cyclophilin A (Zahn, R., Buckle, A. M., Perret, S., Johnson, C. M. J., Corrales, F. J., Golbik, R. & Fersht, A. R. (1996) *Proc. Natl. Acad. Sci. USA* **93**, 15024-15029) are carried out as described.

10

10      **In vivo complementation experiments.** Complementation experiments at 43°C are performed by transforming the thermosensitive (ts) *E. coli* strains SV2 or SV6 (Zeilstra-Ryalls, J., Fayet, O., Baird, L. & Georgopoulos, C. (1993) *J. Bacteriol.* **175**, 1134-1143) with the pJC or pBAD30 series of expression vectors. 3 ml cultures of 15 appropriately transformed cells are grown o/n at 30°C in LB. Cells ( $A_{600} = 0.1$ ) are serially diluted 10-fold in sterile 0.85% (w/v) NaCl. 5 µL from each 10-fold dilution is spotted onto each of two LB plates. One plate is incubated o/n at 37°C and the other overnight at 43°C. The number of viable cells/ml of culture is deduced from the number of colonies in the lowest dilution to give single colonies.

20

20      P1 transduction (35) , using strain AI90 ( $\Delta groEL::kan^R$ ) [pBAD-EL] as donor (Ivic, A., Olden, D., Wallington, E. J. & Lund, P. A. (1997) *Gene* **194**, 1-8), is used to delete the *groEL* gene of TG1 cells transfected by the different pACYC184 or pJC or pBAD30 (sht-GroEL minichaperones) vectors. Transductants are selected on LB plates containing 10 µg/ml of kanamycin at 37°C. Approximately 25 colonies are transferred onto plates containing kanamycin at 50 µg/ml. After incubation for 24 h at 37°C, colonies that grew are screened by PCR as described (Ivic, A., Olden, D., Wallington, E. J. & Lund, P. A. (1997) *Gene* **194**, 1-8). AI90 ( $\Delta groEL::kan^R$ ) [pBAD-EL] cells are transformed with the pJC(sht-GroEL minichaperones) vectors. Transformants are 25 selected at 37°C on LB supplemented with 50 µg/ml of kanamycin, 120 µg/ml of ampicillin, 25 µg/ml of chloramphenicol and 0.2% L(+)arabinose. Depletion of 30

GroEL protein is analysed at 37°C by plating the same quantity of Al90 [pBAD-EL + pJC (sht-GroEL minichaperones)] cells on LB plates containing 1% D(+)glucose or various amount of arabinose.

- 5 Each experiment is performed in triplicate. Plasmids carrying no groE genes or encoding the GroE proteins are used as negative or positive controls, respectively.

**Effect on Lorist6 replication of over-expressing GroE proteins or the minichaperones.** TG1 (Gibson, T. J. (1984) Ph.D. thesis, University of Cambridge)

10 cells carrying the bacteriophage λ origin vector, Lorist6 which encodes kanamycin resistance (Gibson, T. J., Rosenthal, A. & Waterston, R. H. (1987) *Gene* 53, 283-286), are transformed with the pBAD30 series of plasmids. Transformants are selected at 37°C on LB supplemented with 50 µg/ml of kanamycin, 120 µg/ml of ampicillin and 1.0% D(+)glucose. 10-fold serial dilutions of overnight cultures are plated onto 15 kanamycin + ampicillin LB with or without 0.2% L(+)arabinose. The effect of over-expressing GroE proteins or minichaperones on Lorist6 replication is determined by comparing the number of colonies forming units (cfu) per ml resistant to kanamycin on plates containing arabinose relative to the number formed on plates lacking arabinose at 37°C.

20

### **Example 1**

#### **Preparation of chaperone fragments and analysis of *in vitro* folding activity**

*Identification of minimal GroEL chaperone fragment.* The crystal structure of the

25 active minichaperone GroEL(191-345), solved at 2.5 Å, displays a well-ordered domain with the same fold as in intact GroEL chaperonin (Zahn, R., Buckle, A. M., Perret, S., Johnson, C. M. J., Corrales, F. J., Golbik, R. & Fersht, A. R. (1996) *Proc. Natl. Acad. Sci. USA* 93, 15024-15029) (Fig. 2). No electron density is observed for the C-terminal residues 337-345, corresponding to the first half of α-helix H11. Residue 30 Gly335 is at the end of a β-strand. Further truncation before residue 329 leads to inclusion-body formation indicating considerable destabilisation of the over-expressed

GroEL fragment. Residue Glu191 protrudes from the protein surface, making no interactions. Gly192 NH C-caps  $\alpha$ -helix H12 which is absent in the functional monomeric minichaperone GroE(191-345). The N-terminal  $\beta$ -strand starts at Met193 in the hydrophobic core. The integrity of the hydrophobic core is considered to 5 stabilise the polypeptide-binding site composed by  $\alpha$ -helices H8 and H9 and the surrounding loops (Buckle, A. M., Zahn, R. & Fersht, A. R. (1997) *Proc. Natl. Acad. Sci. USA* **94**, 3571-3575; Fenton, W. A., Kashi, Y., Furtak, K. & Horwich, A. L. (1994) *Nature* **371**, 614-619) . On structural grounds, it is predicted that GroEL(193-335) should be the minimal active chaperone unit (Fig. 2).

10

Various N- and C-terminally truncated fragments of the apical domain of GroEL are amplified and cloned downstream of the T7 promoter of pRSETAsht-Eag I vector. The minichaperones sht-GroEL(191-345) and sht-GroEL(191-376) have previously been expressed with a short histidine tail ("sht") for ease of purification (Zahn, R., Buckle, 15 A. M., Perret, S., Johnson, C. M. J., Corrales, F. J., Golbik, R. & Fersht, A. R. (1996) *Proc. Natl. Acad. Sci. USA* **93**, 15024-15029). After sonication, the soluble fractions of IPTG-induced transfected C41(DE3) cells are analysed by SDS-PAGE. sht-GroEL(193-335) is over-expressed in C41(DE3) cells (Miroux, B. & Walker, J. E. (1996) *J. Mol. Biol.* **260**, 289-298) to give ~ 150 mg purified protein per L of culture. 20 Further truncations leads to severe aggregation in C41(DE3) cells. Fragments (all with sht) 193-335, 193-336, 193-337, 193-345, 191-335, 191-336, 191-337 and 191-345 are all expressed at 100-150 mg/L culture after purification and are highly soluble. 193-330, 195-330 and 195-335 are all poorly expressed and have low solubility. sht-GroEL(193-335) purified by gel-filtration is monomeric at  $\mu$ M-mM concentrations, as 25 determined by light scattering and NMR experiments.

The minichaperone GroEL(193-335), lacking the N-terminal histidine-tail, is generated by thrombin cleavage of purified sht-GroEL(193-335). The circular dichroism spectra of GroEL(193-335) with or without the short histidine tail indicated significant  $\alpha$ -helical structure. sht-GroEL(193-335) has been found by differential scanning calorimetry experiments to be the most thermostable of the minichaperones with a  $T_m$  30

at about 70°C. The calorimetric data are also consistent with the unfolding of the minichaperone sht-GroEL(193-335) as a monomer.

*In vitro* activity of minichaperone GroEL(193-335). *In vitro*, rhodanese refolds in high  
5 yield only in the presence of GroEL, ATP, and the co-chaperonin GroES (Zahn, R.,  
Perrett, S., Stenberg, G. & Fersht, A. R. (1996) *Science* 271, 642-645).  
Minichaperone sht-GroEL(193-335) is as active *in vitro* as sht-GroEL(191-345) and  
more active than sht-GroEL(191-376) in chaperoning the folding of rhodanese (Fig. 3).  
However, all three chaperone fragments retain *in vitro* activity.

10

The presence of the N-terminal histidine-tail does not abolish binding activity.

Residue Y203 is in the polypeptide-binding site of GroEL(191-376) minichaperone  
(Fig. 2) (Buckle, A. M., Zahn, R. & Fersht, A. R. (1997) *Proc. Natl. Acad. Sci. USA*  
15 **94**, 3571-3575) . The mutation Y203E prevents the binding of intact GroEL to the  
human ornithine transcarbamylase (Fenton, W. A., Kashi, Y., Furtak, K. & Horwich,  
A. L. (1994) *Nature* 371, 614-619) . The mutation Y203E similarly abolishes the  
chaperone activity of sht-GroEL(193-335) (Fig. 3). This suggests that the recognition  
of denatured polypeptide substrates by GroEL and minichaperones uses the same  
20 residues.

## Example 2

### *In vivo* complementation by chaperone fragments of *ts* chaperone mutants

25 To determine whether GroEL minichaperones can supplement defective GroEL in  
general cell growth, complementation of two thermosensitive (*ts*) *groEL* mutants of *E.*  
*coli* at 43°C is examined (Fig. 4). *E. coli* SV2 has the mutation Glu191→Gly in  
GroEL corresponding to *groEL44* allele , while SV6 carries the *EL673* allele, which  
has two mutations, Gly173→Asp and Gly337→Asp(Zeilstra-Ryalls, J., Fayet, O.,  
30 Baird, L. & Georgopoulos, C. (1993) *J. Bacteriol.* 175, 1134-1143). The

thermosensitivity at 43°C of the SV2 and SV6 strains is suppressed by transformation of pJCGroESL containing the *groE* operon or by GroEL alone (Fig. 4).

Only the minichaperone sht-GroEL(193-335) complements the defect in SV2. About 5 45% of the cells, transformed by the vector encoding the minichaperone sht-GroEL(193-335) which grow at the permissive temperature of 37°C, also grow at 43°C (Fig. 4). In contrast, the defective *groEL* in SV6 is complemented by expression of the minichaperone sht-GroEL(191-345), and less well by sht-GroEL(193-335) (see Fig 4). About 65% of the cells SV6 transformed with pJCsht-GroEL(191-345), which grow at 10 37°C, also grow at 43°C (Fig. 4). Colony-forming units are not observed for either strain at 43°C with vectors either lacking inserts or containing the mutation, Y203E, in the minichaperone gene. Thus, the mutation Y203E, which prevents the growth of LG6 *E. coli* strain (Fenton, W. A., Kashi, Y., Furtak, K. & Horwich, A. L. (1994) *Nature* 371, 614-619), also inactivates the minichaperones *in vivo*. Similar results are 15 obtained using the pBAD30(sht-GroELminichaperone) expression system.

### Example 3

#### *In vivo* complementation at 37°C

20 The effects of minichaperones on the growth at 37°C of a strain of *E. coli* in which the chromosomal *groEL* gene has been deleted are analysed in two ways. First, the *groEL* gene of TG1 which had been transformed with the different pJC or pBAD30 (sht-GroEL minichaperone) vectors is deleted. This is done using P1 transduction from the strain AI90 where the chromosomal *groEL* gene has been precisely replaced by a 25 kanamycin resistance (*kan*<sup>R</sup>) cassette (Ivic, A., Olden, D., Wallington, E. J. & Lund, P. A. (1997) *Gene* 194, 1-8). However, no *kan*<sup>R</sup> transductants can be obtained where the *groEL* gene had been deleted, unless intact GroEL is expressed from the complementing plasmid. This is consistent with the known essential role of GroEL (Fayet, O., Ziegelhoffer, T. & Georgopoulos, C. (1989) *J. Bacteriol.* 171, 1379-1385).

In a second series of experiments, the complementation of AI90 ( $\Delta groEL::kan^R$ ) [pBAD-EL] *E. coli* strain (Ivic, A., Olden, D., Wallington, E. J. & Lund, P. A. (1997) *Gene* 194, 1-8) is analysed. In this strain, the chromosomal *groEL* gene has been deleted and GroEL is expressed exclusively from a plasmid-borne copy of the gene which can be tightly regulated by the arabinose *PBAD* promoter and its regulatory gene, *araC* (Guzman, L.-M., Belin, D., Carson, M. J. & Beckwith, J. (1995) *J. Bacteriol.* 177, 4121-4130). AraC protein acts as either a repressor or an activator depending on the carbon source used. *PBAD* is activated by arabinose but repressed by glucose. The AI90[pBAD-EL] cells can not grow on medium supplemented with glucose at 37°C. None of the minichaperones are able to suppress this *groEL* growth defect.

We then determine whether the constructs can supplement low levels of GroEL from transfected AI90[pBAD-EL] cells by progressively switching on the plasmid-borne *groEL* gene by increasing arabinose from 0 to 0.2% (Fig. 5). In the absence of arabinose, the only cells that form colonies are those carrying pJC expressing GroEL. About 80% of the cells form colonies compared to the number produced in the presence of 0.2% arabinose, at which concentration all cells are fully induced (100% viability). At 0.01% arabinose, cells transfected with pJC expressing sht alone, sht-GroEL(191-376), sht-GroEL(191-345) or sht-GroEL(193-335)(Y203E) show little colony forming ability. Those containing pJC[sht-GroEL(193-335)], however, produced 15-20% of the number produced in the presence of 0.2% arabinose. At 0.07% arabinose, cells containing pJCGroEL produce as many colonies as those fully induced by 0.2%; pJCsht-GroEL[193-335] 75-80%; pJCsht-GroEL[191-345]; 30-40%; pJCsht-GroEL[191-376] ~20%; and the two controls, pJCsht-GroEL[193-335(Y203E)] and pJCsht, ~15%. Thus, pJCsht-GroEL[193-335] can significantly supplement depleted levels of GroEL.

#### Example 4

30 Effect on  $\lambda$  replication of over-expressing GroEL or the minichaperones.

Heat induction of the *groE* operon has been shown to decrease burst size of  $\lambda$  bacteriophage in *E. coli* (Wegrzyn, A., Wegrzyn, G. & Taylor, K. (1996) *Virology* **217**, 594-597). In contrast, we have found that the over-expression of GroEL alone prevents plaque formation by  $\lambda$  in wild-type strains including SV1. This effect is 5 specific since neither over-expression of GroES alone or together with GroEL causes a significant drop in number of plaques formed. No effects on plaque counts from over-expressing the different minichaperones are found. Although the *groE* operon is named for its effects on the E protein of  $\lambda$  (Georgopoulos, C., Hendrix, R. W., Casjens, S. R. & Kaiser, A. D. (1973) *J. Mol. Biol.* **76**, 45-60), it seems that the main 10 effect of GroEL over-expression is mediated through the  $\lambda$  origin.

Both GroEL and the minichaperones inhibit replication of the Lorist6 plasmid which uses the bacteriophage  $\lambda$  origin. A series of cultures of TG1 carrying Lorist6, encoding kanamycin resistance, and each of the pBAD30 vectors are titered on LB 15 plates containing kanamycin and ampicillin in presence or absence of 0.2% arabinose, which induces expression from the *P<sub>BAD</sub>* promoter (Fig. 6). The percentage of cells forming kanamycin resistant colonies in presence of arabinose compared to the number formed in the absence of arabinose (100%) is shown for each of the expression vectors. Loss of kanamycin resistance reflects inhibition of Lorist6 replication. Over-expression 20 of GroES alone has no effect, while GroEL decrease the number of kanamycin resistant colonies 200-fold (Fig. 6). Each of the minichaperones inhibits Lorist6 replication; kanamycin resistant colonies are decreased by 10- to 23-fold (Fig. 6).

## SEQUENCE LISTING

## (1) GENERAL INFORMATION:

5

## (i) APPLICANT:

- (A) NAME: MEDICAL RESEARCH COUNCIL
- (B) STREET: 20 PARK CRESCENT
- (C) CITY: LONDON
- 10 (E) COUNTRY: UK
- (F) POSTAL CODE (ZIP): W1N 4AL

## (ii) TITLE OF INVENTION: CHAPERONE FRAGMENTS

15

## (iii) NUMBER OF SEQUENCES: 10

## (iv) COMPUTER READABLE FORM:

- (A) MEDIUM TYPE: Floppy disk
- (B) COMPUTER: IBM PC compatible
- 20 (C) OPERATING SYSTEM: PC-DOS/MS-DOS
- (D) SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)

## (2) INFORMATION FOR SEQ ID NO: 1:

25

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 32 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- 30 (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: other nucleic acid

- (A) DESCRIPTION: /desc = "SYNTHETIC PRIMER"

35 (iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

5

ATTCAATATGA ATATTCGTCC ATTGCATGAT CG

32

(2) INFORMATION FOR SEQ ID NO: 2:

10

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 46 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

15

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: /desc = "SYNTHETIC PRIMER"

(iii) HYPOTHETICAL: NO

20

(iv) ANTI-SENSE: NO

25

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

AACGGCCGTT AATTAAGGTG CACCGAAAGA TTTATCCAGA ACTACG

46

30 (2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 36 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

35

(ii) MOLECULE TYPE: other nucleic acid

48

(A) DESCRIPTION: /desc = "SYNTHETIC PRIMER"

(iii) HYPOTHETICAL: NO

5 (iv) ANTI-SENSE: NO

10 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

TAGCTGCCAT ATGGCAGCTA AAGACGTAAA ATTCGG

36

(2) INFORMATION FOR SEQ ID NO: 4:

15

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 36 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

20

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: /desc = "SYNTHETIC PRIMER"

25

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

30

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

ATGTAACGGC CGTTACATCA TGCCGCCCAT GCCACC

36

35

(2) INFORMATION FOR SEQ ID NO: 5:

(i) SEQUENCE CHARACTERISTICS:

49

- (A) LENGTH: 16 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

5

- (ii) MOLECULE TYPE: other nucleic acid
  - (A) DESCRIPTION: /desc = "SYNTHETIC PRIMER"

- (iii) HYPOTHETICAL: NO

10

- (iv) ANTI-SENSE: NO

15

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

16

AATTCAACGG CCGTTA

20 (2) INFORMATION FOR SEQ ID NO: 6:

- (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 16 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

25

- (ii) MOLECULE TYPE: other nucleic acid

- (A) DESCRIPTION: /desc = "SYNTHETIC PRIMER"

30

- (iii) HYPOTHETICAL: NO

- (iv) ANTI-SENSE: NO

35

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

50

AGCTTAACGG CCGTTG

16

## (2) INFORMATION FOR SEQ ID NO: 7:

5

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 17 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single

10 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

15

(iv) ANTI-SENSE: NO

20

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

Met Arg Gly Ser His His His His His Gly Leu Val Pro Arg Gly

1 5 10 15

25

Ser

## (2) INFORMATION FOR SEQ ID NO: 8:

30

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 15 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single

35 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: /desc = "SYNTHETIC PRIMER"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

10

15

ATTCATCAAC AAGCC

(2) INFORMATION FOR SEQ ID NO: 9:

15

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 17 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

20

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: /desc = "SYNTHETIC PRIMER"

(iii) HYPOTHETICAL: NO

25

(iv) ANTI-SENSE: NO

30

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

17

TCAGGAGACA GGTAGCC

35 (2) INFORMATION FOR SEQ ID NO: 10:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 548 amino acids

52

- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

5 (ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

10

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

15

Met Ala Ala Lys Asp Val Lys Phe Gly Asn Asp Ala Arg Val Lys Met  
1 5 10 15

20

Leu Arg Gly Val Asn Val Leu Ala Asp Ala Val Lys Val Thr Leu Gly  
20 25 30

25

Pro Lys Gly Arg Asn Val Val Leu Asp Lys Ser Phe Gly Ala Pro Thr  
35 40 45

30

Ile Thr Lys Asp Gly Val Ser Val Ala Arg Glu Ile Glu Leu Glu Asp  
50 55 60

35

Lys Phe Glu Asn Met Gly Ala Gln Met Val Lys Glu Val Ala Ser Lys  
65 70 75 80

Ala Asn Asp Ala Ala Gly Asp Gly Thr Thr Thr Ala Thr Val Leu Ala  
85 90 95

Gln Ala Ile Ile Thr Glu Gly Leu Lys Ala Val Ala Ala Gly Met Asn  
100 105 110

Pro Met Asp Leu Lys Arg Gly Ile Asp Lys Ala Val Thr Ala Ala Val  
115 120 125

53

Glu Glu Leu Lys Ala Leu Ser Val Pro Cys Ser Asp Ser Lys Ala Ile  
130 135 140

5 Ala Gln Val Gly Thr Ile Ser Ala Asn Ser Asp Glu Thr Val Gly Lys  
145 150 155 160

Leu Ile Ala Glu Ala Met Asp Lys Val Gly Lys Glu Gly Val Ile Thr  
165 170 175

10 Val Glu Asp Gly Thr Gly Leu Gln Asp Glu Leu Asp Val Val Glu Gly  
180 185 190

Met Gln Phe Asp Arg Gly Tyr Leu Ser Pro Tyr Phe Ile Asn Lys Pro  
15 195 200 205

Glu Thr Gly Ala Val Glu Leu Glu Ser Pro Phe Ile Leu Leu Ala Asp  
210 215 220

Lys Lys Ile Ser Asn Ile Arg Glu Met Leu Pro Val Leu Glu Ala Val  
20 225 230 235 240

Ala Lys Ala Gly Lys Pro Leu Leu Ile Ile Ala Glu Asp Val Glu Gly  
245 250 255

25 Glu Ala Leu Ala Thr Leu Val Val Asn Thr Met Arg Gly Ile Val Lys  
260 265 270

Val Ala Ala Val Lys Ala Pro Gly Phe Gly Asp Arg Arg Lys Ala Met  
30 275 280 285

Leu Gln Asp Ile Ala Thr Leu Thr Gly Gly Thr Val Ile Ser Glu Glu  
290 295 300

Ile Gly Met Glu Leu Glu Lys Ala Thr Leu Glu Asp Leu Gly Gln Ala  
35 305 310 315 320

Lys Arg Val Val Ile Asn Lys Asp Thr Thr Thr Ile Ile Asp Gly Val

54  
325                   330                   335  
Gly Glu Glu Ala Ala Ile Gln Gly Arg Val Ala Gln Ile Arg Gln Gln  
340                   345                   350  
Ile Glu Glu Ala Thr Ser Asp Tyr Asp Arg Glu Lys Leu Gln Glu Arg  
355                   360                   365  
Val Ala Lys Leu Ala Gly Gly Val Ala Val Ile Lys Val Gly Ala Ala  
10                   370                   375                   380  
Thr Glu Val Glu Met Lys Glu Lys Lys Ala Arg Val Glu Asp Ala Leu  
385                   390                   395                   400  
His Ala Thr Arg Ala Ala Val Glu Glu Gly Val Val Ala Gly Gly Gly  
15                   405                   410                   415  
Val Ala Leu Ile Arg Val Ala Ser Lys Leu Ala Asp Leu Arg Gly Gln  
420                   425                   430  
Asn Glu Asp Gln Asn Val Gly Ile Lys Val Ala Leu Arg Ala Met Glu  
20                   435                   440                   445  
Ala Pro Leu Arg Gln Ile Val Leu Asn Cys Gly Glu Glu Pro Ser Val  
450                   455                   460  
Val Ala Asn Thr Val Lys Gly Gly Asp Gly Asn Tyr Gly Tyr Asn Ala  
25                   465                   470                   475                   480  
Ala Thr Glu Glu Tyr Gly Asn Met Ile Asp Met Gly Ile Leu Asp Pro  
30                   485                   490                   495  
Thr Lys Val Thr Arg Ser Ala Leu Gln Tyr Ala Ala Ser Val Ala Gly  
500                   505                   510  
Leu Met Ile Thr Thr Glu Cys Met Val Thr Asp Leu Pro Lys Asn Asp  
35                   515                   520                   525

55

Ala Ala Asp Leu Gly Ala Ala Gly Gly Met Gly Met Gly Met  
530 535 540

Gly Gly Met Met

5 545

Claims

1. A method for determining whether a fragment of a molecular chaperone is active *in vivo*, comprising the steps of:
  - a) providing a cell with a deficient molecular chaperone activity;
  - b) administering the molecular chaperone fragment to the cell;
  - c) determining whether the molecular chaperone fragment complements the deficient endogenous molecular chaperone activity.
- 10 2. A method according to claim 1, wherein the molecular chaperone activity in the cell is temperature sensitive such that, when the cell is grown at a non-permissive temperature, the molecular chaperone activity is deficient.
- 15 3. A method according to claim 1, wherein the molecular chaperone activity in the cell is under the control of regulatable control sequences such that, in response to the appropriate stimulus, the molecular chaperone activity can be downregulated such that it is deficient.
- 20 4. A method according to any one of claims 1 to 3, wherein the complementation of the deficient molecular chaperone activity results in improved cell growth.
- 25 5. A method according to any one of claims 1 to 3, wherein the complementation of the deficient molecular chaperone activity results in a detectable signal from a reporter system.
6. A method according to any preceding claim, wherein the molecular chaperone fragment is between 100 and 500 amino acids in length.
- 30 7. A method according to any preceding claim, wherein the chaperone fragment is selected from the group consisting of fragments of the molecular chaperones p90

Calnexin; HSP family; HSP 70 family; DNA K; DNAJ; HSP 60 family; GroEL; ER-associated chaperones; HSP 90; Hsc 70; sHsps; SecA; SecB; Trigger factor; zebrafish hsp 47, 70 and 90; HSP 47; GRP 94; Cpn 10; BiP; GRP 78; C1p, FtsH; Ig invariant chain; mitochondrial hsp 70; EBP; mitochondrial m-AAA; Yeast Ydj1; Hsp 104; 5 ApoE; Syc; Hip; TriC family; CCT; PapD and calmodulin.

8. A method according to claim 7, wherein the chaperone fragment is a fragment of a molecular chaperone of the HSP60 class.

10 9. A method according to claim 8, wherein the chaperone fragment is a fragment of *E. coli* GroEL.

10. A method according to claim 9, wherein the fragment is a fragment of the apical domain of *E. coli* GroEL.

15 11. A method according to claim 10, wherein the fragment is selected from the group consisting of residues 191-376, 191-345 and 193-335 of *E. coli* GroEL.

12. A method according to any preceding claim, wherein the molecular chaperone 20 fragment is administered to the cell by introducing into the cell a nucleic acid encoding the fragment and expressing the nucleic acid in the cell.

13. A method for providing a chaperone activity *in vivo*, comprising administering to a cell a fragment of a molecular chaperone which has *in vivo* chaperone activity.

25 14. A method according to claim 13, wherein the molecular chaperone fragment complements a mutant or depressed endogenous molecular chaperone activity.

15. A method according to claim 13 or claim 13, comprising the steps of:

- a) providing a nucleic acid encoding a fragment of a molecular chaperone having *in vivo* chaperone activity, operably linked to control sequences active in an intended host cell;
- b) inserting the nucleic acid into the intended host cell;
- 5 c) expressing the nucleic acid in the host cell such that the molecular chaperone fragment is produced.

16. Use of a fragment of a molecular chaperone to complement a mutant or deficient molecular chaperone activity *in vivo*.

# INTERNATIONAL SEARCH REPORT

International Application No

PCT/GB 98/02042

**A. CLASSIFICATION OF SUBJECT MATTER**  
IPC 6 G01N33/50 C12Q1/02

According to International Patent Classification (IPC) or to both national classification and IPC

**B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)  
IPC 6 G01N C12Q C07K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

Category <sup>a</sup>	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	EP 0 333 201 A (GEN HOSPITAL CORP) 20 September 1989 see claims 1-6 see column 15, line 28 - column 17, line 21 ---	1-16
P, X	WO 98 15629 A (CZECH MICHAEL P ;KLARLUND JES K (US); UNIV MASSACHUSETTS (US)) 16 April 1998 see claims 41,46-73 see page 3, line 26 - page 4, line 4 see page 8, line 3 - page 9, line 29 see page 17, line 22 - page 18, line 29 see page 48, line 25 - page 51, line 23 see page 56, line 19 - page 60, line 26 --- -/--	1-16

Further documents are listed in the continuation of box C.

Patent family members are listed in annex.

\* Special categories of cited documents :

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search

30 November 1998

Date of mailing of the international search report

07/12/1998

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl.  
Fax: (+31-70) 340-3016

Authorized officer

Routledge, B

# INTERNATIONAL SEARCH REPORT

International Application No

PCT/GB 98/02042

**C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT**

Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P, X	WO 98 13496 A (MEDICAL RES COUNCIL ;FERSHT ALAN ROY (GB); ZAHN RALPH (GB); ALTAMI) 2 April 1998 cited in the application see claims 29-33,47 see page 27, line 5 - page 31, line 26 see page 33, line 8 - line 10 ----	1-16
X	WO 97 14794 A (UNIV DUNDEE ;LANE DAVID PHILIP (GB); HUPP THEODORE ROBERT (GB)) 24 April 1997 see claims 8-13 see page 2, line 29 - line 34 see page 5, line 28 - page 6, line 14 see page 27, line 27 - page 28, line 14 see page 29, line 21 - page 30, line 2 see page 32, line 36 - page 34, line 22 ----	1-16
X	WO 97 10001 A (UNIV FORDHAM) 20 March 1997 see claims see page 15, line 29 - page 18, line 4 see page 29, line 17 - line 24 see page 38, line 31 - page 43, line 13 see page 55, line 5 - page 56, line 21 ----	1-16
X	WO 95 15338 A (UNIV QUEENSLAND ;MORTON HALLE (AU); CAVANAGH ALICE CHRISTINA (AU)) 8 June 1995 see claims 4-10 see page 2, line 5 - line 20 see page 5, line 27 - line 31 see page 6, line 13 - line 21 see page 17, line 32 - page 23, line 27 ----	1-16
E	US 5 780 034 A (COHEN IRUN R ET AL) 14 July 1998 see claims 5-18 see column 4, line 5 - line 35 see examples 10,11,20 see column 24, line 66 - column 26, line 14 -----	1-16

## INTERNATIONAL SEARCH REPORT

International application No.

PCT/GB 98/02042

### Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1.  Claims Nos.: 13-16  
because they relate to subject matter not required to be searched by this Authority, namely:  
Although claims 13-16 are directed to a method of treatment of the human/animal body, the search has been carried out and based on the alleged effects of the compound/composition.  
Claims relate to methods of therapy of the human or animal body
2.  Claims Nos.:  
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:
3.  Claims Nos.:  
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

### Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1.  As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.
2.  As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3.  As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:
4.  No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

#### Remark on Protest

- The additional search fees were accompanied by the applicant's protest.
- No protest accompanied the payment of additional search fees.

**INTERNATIONAL SEARCH REPORT**

Information on patent family members

International Application No	
PCT/GB 98/02042	

Patent document cited in search report	Publication date	Patent family member(s)		Publication date
EP 0333201 A	20-09-1989	AT 112168 T AU 645243 B AU 3280489 A DE 68918479 D DE 68918479 T DK 224790 A ES 2064375 T IE 62447 B JP 3503410 T PT 90019 A,B WO 8908661 A US 5137805 A		15-10-1994 13-01-1994 05-10-1989 03-11-1994 02-03-1995 15-11-1990 01-02-1995 08-02-1995 01-08-1991 20-11-1990 21-09-1989 11-08-1992
WO 9815629 A	16-04-1998	AU 4749497 A		05-05-1998
WO 9813496 A	02-04-1998	WO 9824909 A AU 4467497 A AU 1036397 A		11-06-1998 17-04-1998 29-06-1998
WO 9714794 A	24-04-1997	AU 7317496 A EP 0859839 A		07-05-1997 26-08-1998
WO 9710001 A	20-03-1997	US 5837251 A AU 7018196 A EP 0859631 A		17-11-1998 01-04-1997 26-08-1998
WO 9515338 A	08-06-1995	AU 684577 B AU 1103495 A AU 684578 B AU 1103695 A WO 9515339 A CA 2176948 A CA 2177179 A EP 0731811 A EP 0733071 A JP 9506431 T JP 9505811 T NZ 276670 A NZ 276672 A		18-12-1997 19-06-1995 18-12-1997 19-06-1995 08-06-1995 08-06-1995 08-06-1995 18-09-1996 25-09-1996 24-06-1997 10-06-1997 26-01-1998 26-01-1998
US 5780034 A	14-07-1998	US 5114844 A US 5578303 A AU 639456 B AU 5546790 A CA 2029861 A EP 0417271 A FI 961398 A JP 4502920 T NO 300594 B NO 964483 A US 5671848 A WO 9010449 A		19-05-1992 26-11-1996 29-07-1993 09-10-1990 15-09-1990 20-03-1991 27-03-1996 28-05-1992 23-06-1997 22-10-1996 30-09-1997 20-09-1992